NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

STATISTICAL ANALYSIS OF WARFARE: IDENTIFICATION OF WINNING FACTORS WITH A FOCUS ON IRREGULAR WARFARE

by

Bilal S. Gondal

September 2015

Thesis Advisor: Thomas W. Lucas
Second Reader: Samuel E. Buttrey

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704–0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** September 2015	**3. REPORT TYPE AND DATES COVERED** Master's thesis	
4. TITLE AND SUBTITLE STATISTICAL ANALYSIS OF WARFARE: IDENTIFICATION OF WINNING FACTORS WITH A FOCUS ON IRREGULAR WARFARE			**5. FUNDING NUMBERS**
6. AUTHOR(S) Gondal, Bilal S.			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ____N/A____.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

The purpose of this study was to determine important factors in winning conventional and irregular conflict. The research sought to identify variables and trends for conventional and irregular warfare as a means for predicting battle outcomes. The variables related to conventional and irregular warfare differ. There are limited variables for analysis of irregular conflicts due to the complexity of data collection during these conflicts. Selected variables from both types of conflict were synthesized using a descriptive statistics and decision tree methodology to identify important trends in warfare. The analysis indicated that cavalry, artillery, close air support, air superiority, leadership, and initiative played vital roles in deciding the outcome of conventional battles over time. The exploration of irregular warfare revealed that the population plays a major role in these conflicts. The numbers of participants are higher and the duration is longer in irregular conflict than in conventional warfare. These irregular conflicts primarily occurred in areas of low gross domestic product, low employment-to-population ratio, and government ineffectiveness.

14. SUBJECT TERMS winning wars, irregular warfare, combat modeling, statistical analysis of warfare, causal factors for wars, conflicts			**15. NUMBER OF PAGES** 273
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540–01-280-5500 Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

STATISTICAL ANALYSIS OF WARFARE: IDENTIFICATION OF WINNING FACTORS WITH A FOCUS ON IRREGULAR WARFARE

Bilal S. Gondal
Major, Pakistan Army
B.S., Pakistan Military Academy, 2001

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 2015

Approved by: Thomas W. Lucas
Thesis Advisor

Samuel E. Buttrey
Second Reader

Patricia A. Jacobs
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The purpose of this study was to determine important factors in winning conventional and irregular conflict. The research sought to identify variables and trends for conventional and irregular warfare as a means for predicting battle outcomes. The variables related to conventional and irregular warfare differ. There are limited variables for analysis of irregular conflicts due to the complexity of data collection during these conflicts. Selected variables from both types of conflict were synthesized using a descriptive statistics and decision tree methodology to identify important trends in warfare. The analysis indicated that cavalry, artillery, close air support, air superiority, leadership, and initiative played vital roles in deciding the outcome of conventional battles over time. The exploration of irregular warfare revealed that the population plays a major role in these conflicts. The numbers of participants are higher and the duration is longer in irregular conflict than in conventional warfare. These irregular conflicts primarily occurred in areas of low gross domestic product, low employment-to-population ratio, and government ineffectiveness.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ACLED	Armed Conflict Location Event Dataset
ADV	Advantage
AERO	Relative Air Superiority in Theater
AIC	Akaike Information Criterion
Arty	Artillery Ratio
CAA	Army Concepts Analysis Agency
CAIC	Bozdogan's Consistent Akaike Information Criterion
CEA	Relative Control Effectiveness
CER	Casualty Exchange Ratio
CHASE	Combat History Analysis Study Effort
CI	Counter Intelligence
CIA	Central Intelligence Agency
COIN	Counter Insurgency Operations
COW	Correlate of War Projects
DE	Double Envelopment
DL	Delaying Action Adopted
DMSO	Defense Modeling and Simulation Office
DO	Defensive Plan
DOD	Department of Defense
EE	Single Envelopment
EPS	Bitterness
FATA	Federally Administrated Tribal Areas
FD	Fortified Defense
FER	Fractional Exchange Ratio
FF	Frontal Attack
FLY	Close Air Support Ratio
FR	Force Ratio
FRCI	Republican Forces of Ivory Coast
FUNCINPEC	National United Front for an Independent, Neutral, Peaceful, and Cooperative Cambodia

GDP	Gross Domestic Product
HD	Hasty Defense
HERO	Historical Evaluation and Research Organization
IDPs	Internally Displaced Persons
IFR	Initial Force Ratio
INITA	Relative Imitative Advantage
INTELA	Relative intelligence Advantage
ISAF	International Security Assistance Force
ISIS	Islamic State of Iraq and Syria
IW	Irregular Warfare
JOC	Joint Operating Concept
LEADA	Relative Leadership Advantage
LOGSA	Relative Logistic Advantage
LTTE	Liberation Tigers of Tamil Eelam
MIC	Managing Intra-State Conflict
MILC	Managing Intra-State Low Intensity Conflict
MILOCI	Movement for the Liberation of Western Ivory Coast
MOMNTA	Relative Momentum Advantage
NATO	North Atlantic Treaty Organization
NPFL	National Patriotic Front of Liberia
NSAs	Non-State Armed Groups
NTEs	Non-Territorial Entities
OSS	Office of Strategic Services
PD	Prepared Defense
POST1	Defender Primary Posture
PRIAI	Attackers Primary Tactical Scheme
RC	River Crossing
RESADV	Residual Advantage
RMSE	Root Mean Square Error
SD	Standard Deviation
SOCOM	Special Operations Command
SSTR	Stability Security Transition and Reconstructions

SURPA	Surprise
TANK	Tank Ratio
TECHA	Relative Technology Advantage
TRNGA	Relative Training Advantage
U.S	United States
UCDP	Uppsala Conflict Data Program
UN	United Nations
UNMA	United Nation Mission in Afghanistan
UNMIH	United Nation Mission in Haiti
WD	Withdrawal Adopted
WINA	Winner
WWI	World War I
WWII	World War II

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Understanding the winning of battles is becoming more arduous with the dynamic, ever-evolving nature of warfare. It is critical for military leaders to forecast the future of warfare from the prism of the past. The twentieth century saw a transformation from conventional to irregular warfare in which armies started combating non-state, widely dispersed groups. Although this change appears to be quite simple, it has had a deep impact on military strategy, organization, and the way wars are fought. Non-state actors are challenging the technological, financial, and organizational superiority of well-trained armies by using terrorism as an instrument for perpetuating their agendas. Non-state actors have been successful because many armies around the globe are still primarily focused on fighting conventional wars rather than countering evolving threats.

There has been little work that bridges the gap between qualitative and quantitative analysis in this area. It is this void between the two methods of studying warfare that results in misunderstandings at the executional end. For the most part, warfare has been studied separately in the two spheres. This research aims to merge qualitative and quantitative methods to explain similarities and differences between conventional and irregular warfare. The research used four different datasets to explore trends in both conventional and irregular warfare. For conventional warfare, the CDB90G dataset compiled by the Center for Army Analysis (CAA) was used to identify important trends and associated factors with winning across different time frames. For irregular warfare, the Armed Conflict Location and Event Dataset (ACLED), Correlates of War (COW) Project, and University of Uppsala (UCDP) datasets were used to explore trends along with potentially causal factors.

Statistical techniques can provide military planners detailed insight into distinctive aspects of evolving warfare. The research used both descriptive statistics along with decision tree and boosted tree modeling techniques to determine important variables associated with different kinds of warfare over time. Analysis of the CDB90G dataset revealed that the important variables associated with winning battles have changed over time. In the fifteenth century, leadership, cavalry, force ratio, and initiative were the

major variables for winning battles. In the sixteenth century, leadership, cavalry, artillery, and terrain became the most important variables. In the nineteenth century, which overlaps with the Industrial Revolution, cavalry, leadership, and artillery emerged as the most important variables. Beginning in the twentieth century, conventional warfare became more dependent on modern technologies. Between 1900 and 1920, artillery, force ratio, and initiative became more dominant. Between 1937 and 1945, air superiority, tank ratio, force ratio, initial force ratio, artillery, scheme of attack, and weather gained prominence. Finally, between 1951 and 1982, close air support, air superiority, force ratio, initial force ratio, defensive posture, and weather were the factors most associated with battle outcomes.

The analysis of datasets for irregular warfare revealed that these conflicts are more complicated and more difficult to understand than conventional battles. The variables that differentiate irregular warfare and tend to make it more complicated are the number of participants involved, involvement of civilian populations, foreign government involvement, duration of conflicts, categories of violence, seasonality (weather) effect, geographical location, and the tendency for events to spill outside national borders.

The analysis of irregular warfare identified the following participant trends. The numbers of participants involved in irregular warfare were far greater than those in regular warfare, especially when the conflicts involved foreign elements. Another noticeable pattern was that non–state actors involved in these conflicts operate within relatively small networks. These networks have the ability to conceal themselves within civilian populations, which makes it difficult to identify them. This is in contrast to the conventional battles in which organized armies practice Clausewitz's principle of massing power at one decisive point. These give non-state actors the capability of engaging multiple different fronts simultaneously. This network approach also makes them more resilient to action by militaries. The major difference between irregular and conventional warfare is that in irregular conflicts, civilians, as opposed to military units, are major targets. Non-state actors use civilian populations to perpetuate their agendas, coercing them to support their cause or suffer brutal consequences. Another significant trend that emerged in irregular warfare is participation of foreign governments. Foreign

governments have been involved in supporting local governments directly and indirectly to overcome non-state actors.

Interestingly, the durations of irregular conflicts, compared to conventional ones, tend to be longer. This is due to the ability of non-state actors to prolong war without burdening economies. In the case of conventional warfare, countries cannot afford to prolong conflicts because it drains their economy considerably. Another trend is that in irregular warfare, the types of events are quite different from conventional battles: instead of mounting offensive operations on opposing armies, most of the events that take place are battles with no change of territory. Moreover, in irregular conflicts, non-state actors tend to increase violence to destabilize the government. An analysis of irregular conflicts also revealed that weather is a dominant factor used by non-state actors to support their campaigns against governments. It is common for non-state actors to launch their campaigns in moderate weather and essentially hibernate during extreme conditions.

From the data on irregular warfare, research identified geography as playing a vital role in irregular conflicts. Non-state actors tend to establish their footholds within remote territories. Non-state actors mostly use mountains, ravines, and harsh terrain to their advantage. Moreover, the use of cross-border movement by non-state actors from a country where conflict is occurring into a neighboring country usually takes advantage of unguarded borders. The analysis of data revealed that irregular conflicts tend to spill outside the country of origin and, in most cases, lead to events in neighboring countries.

Out of all the regions analyzed, Africa has exhibited an exponential growth of violence in the last few years. However, the dataset was missing information on other regions of extreme violence, including Afghanistan and Iraq. If these conflicts are not addressed, Africa has the potential to become the future hotspot where non-state actors and terrorist organizations can train and flourish. The research was extended to gain insight into factors that cause these conflicts using a boosted tree method. The analysis revealed that poor political stability, low gross domestic product, low population to employment ratio, and government ineffectiveness are some of the factors associated with these conflicts.

This research is a first step in uncovering some of the intricacies associated with warfare. The subject is quite vast, so this project focused on just a few aspects. There are many aspects related to warfare, especially to irregular warfare, that need exploration. The first and most important thing for gaining more insight is through formulating a database from a military perspective to analyze these conflicts. This can give us more detailed insight and help future military leaders to plan based on mathematical analysis. The second thing that needs more exploration is causal factors for irregular conflicts. By identifying and addressing these factors, world leadership may be able to considerably reduce these conflicts in the future.

ACKNOWLEDGMENTS

I am grateful for Professor Thomas W. Lucas for providing detailed guidance on the subject and walking me through the path of statistical and combat analysis. I am also indebted to Professor Samuel Buttrey for his valuable time during research to figure out complex problems involving datasets. It would be unfair if I don't acknowledge support provided by Noel Yucuis of the Graduate Writing Center, who meticulously helped me to improve the quality and content of the thesis. I am also grateful to my wife, Javaria; son, Mekaeel; and daughter, Ashalina for their complete support during my research. Their smiles made my work easier.

I would like to dedicate this thesis to all those souls who lost their lives for their motherlands. There is no doubt regarding their commitment to the cause they were fighting for.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

"We make war that we may live in peace."

–Aristotle

A. GENERAL INTRODUCTION

Armed conflicts develop when there are disagreements. Disagreements can be over conflicting domains or over a wide spectrum of interests including power mêlées, territorial control, financial interests, politics, and ideology.[1] Disagreements can result in armed conflict if unresolved.[2] Armed conflicts can be categorized into conventional, unconventional, and irregular warfare. The oldest forms of conflicts were irregular in nature. Military analysts also classify warfare into four generations; the first, second, and third generations fall into the category of conventional warfare, while the fourth generation is considered irregular warfare.[3] The classification of warfare into overlapping generations seems to correspond with the evolution of human society.

The Romans marked the transition from irregular to conventional warfare by introducing the concept of the formal army.[4] The Romans introduced military formations and organized the army along modern lines to fight national adversaries. The Romans' military model was neglected "in the west," according to Strong and Gillum, by the European Middle Ages, when battles were mostly confined within the realm of small skirmishes, raids, ambushes, and unconventional techniques.[5]

---

[1] Morton Deutsch, Coleman T Peter, and Marcus C Eric, *The Handbook of Conflict Resolution: Theory and Practice* (Hoboken, NJ: John Wiley & Sons, 2011).

[2] Thomas X Hammes, "The Evolution of War: The Fourth Generation," *Marine Corps Gazette* 78, no. 9 (1994): 35–44.

[3] Ibid.

[4] Carol Strong and Joshua Gillum, "Unraveling the Gordian Knot of Strategic U.S. Military Engagement: Asymmetrical Relationships, Unconventional Means and International Conflict," Paper prepared for Cultural Diplomacy and International Relations: New Actors, New Initiatives and New Targets Institute for Cultural Diplomacy, Annual Conference, Berlin, Germany 15–18 November, 2011.

[5] Ibid.

However, during the seventeenth century, warfare again witnessed a shift from irregular to conventional following the Roman style.[6] This transition from irregular to conventional warfare resulted in the first generation of warfare. This was characterized by formation of formal armies, line-and-column tactics, and identification between civilians and soldiers.[7] The major reason for the shift from irregular to conventional warfare was the development of nation-states.[8] The development of nation-states shifted the power dynamics from small tribal structures to well-defined national systems enforcing taxation and conscription.[9] The concept was further consolidated and developed with transition into the Industrial Revolution of the late 19th century.[10] The development of machine guns and indirect firing weapons became the catalyst for the second generation. The newly developed machine guns and artillery pieces were more lethal than the weapons used in previous wars. These inventions changed the pace of warfare and achieved attrition of enemy forces in limited time by massing firepower.[11] However, along with advancements in technology, warfare continued to transform. The development of armored vehicles moved the world into the third generation of warfare. Speed, surprise, maneuver, attrition, and physical dislocation to collapse the enemy forces characterize the third generation of warfare, commonly known as Blitzkrieg. Military order also witnessed a shift from imposed discipline to self-discipline.[12]

As warfare evolved through the first three generations, it became more civilized with greater respect for adversaries, civilians, and infrastructure. The armies emerged as the primary instrument of national security. War between nations resulted in evolution of modern just war theory based on the philosophy of Saint Augustine. Just war theory has

---

6 Ibid.

7 William S Lind, "Understanding Fourth Generation Warfare," *Military Review* no. 5 (2004): 12–16.

8 Michael Howard, "War and the Nation-State," *Daedalus* 108, no.4 (1979): 101–110.

9 John Frederick Charles Fuller, *The Conduct of War, 1789–1961: A Study of the Impact of the French, Industrial, and Russian Revolutions on War and its Conduct* (Boston, MA: Da Capo Press, 1992).

10 Ibid.

11 William S Lind, "Understanding Fourth Generation Warfare," *Military Review* no. 5 (2004): 12–16.

12 Thomas X Hammes, "War Evolves into the Fourth Generation," *Contemporary Security Policy* 26, no. 2 (2005): 189–221.

three parts: *jus ad bellum*, justice before resorting to war; *jus in bello*, justice during war; and *jus post bellum*, justice at termination of war.[13]

The fourth generation of warfare differs from the first three generations in that it uses irregular means to achieve victory over superior forces.[14] The fourth generation of warfare has challenged the military structure established by the first three generations and has marked the transformation from conventional back to irregular warfare. Irregular warfare is considered an effective instrument against a superior force and challenges hegemony by using pre-modern tactics.[15] The focus of this kind of warfare is to use unpredictable, unconventional means against a superior power to weaken its will to fight.

The properties that differentiate irregular from conventional warfare effectively "blur the lines" between state and non-state actors, soldiers and civilians, politics and war, and peace and conflict zones.[16] The population is considered the vital instrument for waging irregular warfare. Individual citizens are both victims and have the potential to become combatants. There are no defined boundaries for the battlefield; all population centers have the chance to become battlegrounds. The al-Qaeda attack on New York's Twin Towers; the Taliban's attack on the school in Peshawar, Pakistan; and the kidnapping of innocent schoolgirls by Boko Haram in Nigeria are examples of irregular warfare's undefined boundaries, and indiscriminate violence against civilians. In these events the initiative is in the hands of the perpetrators.

In the twenty-first century, there have been very few conventional conflicts fought by Western powers. The major conflicts in which Western powers remain entangled are irregular in nature and fought outside the West. The adversaries in these types of conflicts are non-state actors with their own ulterior motives. Recently, the major actors in these conflicts have been radical Islamic organizations that question the legitimacy of their own governments and the West at the same time. The religious legitimization of the conflicts has made them a complex phenomenon difficult to defeat. Strong and Gillum assert that

---

13 Brian Orend, *The Morality of War* (Calgary, Canada: Broadview Press, 2013).

14 William S Lind, "Understanding Fourth Generation Warfare," *Military Review* no. 5 (2004): 12–16.

15 Ibid.

16 Ibid.

the distinct characteristics of irregular warfare are not easily transferrable to the dialogues associated with conventional warfare.[17] The patterns of irregular warfare are so diverse that strategy and tactics used to win conventional warfare are not applicable. In the succeeding paragraphs, theoretically established definitions of conventional, unconventional, and irregular warfare are discussed to develop a foundation for this research.

1. Conventional War

Conventional warfare is considered the most well-established and customary form of conflict between states. Conventional war, as previously discussed, is often fought with civilized norms, following rules set by the international community. The main focus in conventional war is to defeat the enemy force while ensuring a minimum loss to the opponent's population. The U.S. Army describes conventional warfare as:

> A form of warfare between states that employs direct military confrontation to defeat an adversary's armed forces, destroy an adversary's war-making capacity, or seize or retain territory in order to force a change in an adversary's government or policies. The focus of conventional military operations is normally an adversary's armed forces with the objective of influencing the adversary's government. It generally assumes that the indigenous populations within the operational area are non-belligerents and will accept whatever political outcome the belligerent governments impose, arbitrate, or negotiate. A fundamental military objective in conventional military operations is to minimize civilian interference in those operations.[18]

The U.S. Army's definition of unconventional warfare serves as a guideline for the purpose of this research.

In conventional warfare, the opposing armies have similarities in organization and follow approximately the same means to achieve the desired end state. In the present environment, it is difficult to geographically subjugate an opponent through conventional

---

[17] Carol Strong and Joshua Gillum, "Unraveling the Gordian Knot of Strategic U.S. Military Engagement: Asymmetrical Relationships, Unconventional Means and International Conflict," Paper prepared for Cultural Diplomacy and International Relations: New Actors, New Initiatives and New Targets Institute for Cultural Diplomacy, Annual Conference, Berlin, Germany 15–18 November, 2011.

[18] Department of the Army, *Army Special Operations Forces Unconventional Warfare* (FM 3–05.130) (Washington, D.C: Army Publishing Directorate, 2008).

war. The forced occupation or unlawful violation of the border by an aggressor may be resisted by instruments of global power such as the United Nations, and regional organizations such as the North Atlantic Treaty Organization (NATO). The most distinct character of conventional warfare is that it is fought between states and there is little or no involvement of non-state actors.

2. Unconventional Warfare

The definition of unconventional warfare has been developed over a long period due to continuous developments in warfare.[19] According to the U.S. Army, the creation of the Office of Strategic Services (OSS) during World War II (WWII) was the first step towards unconventional warfare. Initially, Unconventional operations were considered guerrilla and covert operations behind the enemy lines.[20] The first term the U.S. Army used to describe unconventional warfare was "partisan warfare."[21]

According to the U.S. Army, unconventional warfare is composed of escape, evasion, and subversion activities.[22] Over the passage of time, the definition of unconventional warfare evolved to incorporate necessary changes based on operational events. The definition of unconventional warfare is:

> Operations conducted by, with, or through irregular forces in support of a resistance movement, an insurgency, or conventional military operations.[23]

The previously-mentioned definition is considered as the definition of unconventional warfare in this research. Unconventional and irregular warfare tends to overlap and often perplexes military analysts. However, with the development of the

---

[19] Ibid.

[20] Department of the Army, *Special Forces Operations* (FM 31 -20) (Washington, D.C: Army Publishing Directorate, 2001).

[21] Department of the Army, *Army Special Operations Forces Unconventional Warfare* (FM 3–05.130) (Washington, D.C: Army Publishing Directorate, 2008).

[22] Department of the Army, *Special Forces Operations* (FM 31 -20) (Washington, D.C: Army Publishing Directorate, 2001).

[23] Department of the Army, *Army Special Operations Forces Unconventional Warfare* (FM 3–05.130) (Washington, D.C: Army Publishing Directorate, 2008).

irregular warfare definition, unconventional warfare has been added as an activity of irregular warfare.

3. Irregular Warfare

Defining irregular warfare is a difficult process, as it encompasses a wide variety of events and activities. Summarizing all of the events and activities into one definition has remained a gigantic task for military analysts and historians. The draft Irregular Warfare Joint Operating Concept (JOC) acknowledged these difficulties in forming a definition for irregular warfare.[24] These definitional difficulties in language were then used in the September 2007 release of the Irregular Warfare JOC:[25]

> Irregular Warfare is a complex, messy, and ambiguous social phenomenon that does not lend itself to clean, neat, concise, or precise definition. This JOC uses the term in two contexts. First, Irregular Warfare is a form of armed conflict. As such, it replaces the term "low-intensity conflict." Second, Irregular Warfare is a form of warfare. As such, it encompasses insurgency, counterinsurgency, terrorism, and counterterrorism, raising them above the perception that they are somehow a lesser form of conflict below the threshold of warfare.[26]

Rand Corporation's monograph "Assessing Irregular Warfare" claims that the U.S. Special Operations Command (SOCOM) and the Office of the Assistant Secretary of Defense modified the current official definition of irregular warfare from one that first emerged during a workshop hosted for Special Operations and Low Intensity Conflict. The definition was finally approved by Deputy Secretary of Defense Gordon England on April 17, 2006.

> A form of warfare that has as its objective the credibility and or legitimacy of the relevant political authority with the goal of undermining or supporting that authority. Irregular warfare favors indirect approaches, though it may apply the full range of military and other capabilities to seek

---

[24] Eric V Larson, Derek Eaton, Nichiporuk Brian, and Szayna S Thomas, *Assessing Irregular Warfare: A Framework for Intelligence Analysis* (Santa Monica, CA: Rand Corporation, 2009).

[25] Ibid.

[26] Ibid.

> asymmetric approaches, in order to erode an adversary's power, influence and will.[27]

According to the monograph, this definition has been widely used. However, in October 2006, another definition of irregular warfare was introduced, which is now widely accepted in the Department of Defense (DOD) as the definition of irregular warfare. The definition is:

> A violent struggle among state and non-state actors for legitimacy and influence over the relevant populations. Irregular Warfare favors indirect and asymmetric approaches, though it may employ the full range of military and other capacities, in order to erode an adversary's power, influence, and will. It is inherently a protracted struggle that will test the resolve of our Nation and our strategic partners.[28]

These definitions are somewhat different and are unable to fully eliminate the vagueness associated with defining Irregular Warfare. However, they share one commonality: the operating environment associated with success in conventional warfare is not applicable in irregular warfare. Second, success is dependent on winning population support and allegiance rather than on simply defeating the enemy forces. Both definitions emphasize that credibility, legitimacy, and will are the central focuses in irregular warfare. The political concept of gaining the sympathy of the population along with mobilization in favor of the government is largely dependent on indirect and non-military means. The military approach alone in irregular warfare will not result in success.

4. Activities of Irregular Warfare

Irregular Warfare is innately a long, drawn-out conflict that tests the resolve of the defense forces and nation simultaneously.[29] Irregular warfare is also dependent upon demography, culture, religion, and the social outlook of society. Irregular warfare is a violent struggle, but success is not merely dependent on weaponry, platforms, and

---

[27] Eric V Larson, Derek Eaton, Nichiporuk Brian, and Szayna S Thomas, *Assessing Irregular Warfare: A Framework for Intelligence Analysis* (Santa Monica, CA: Rand Corporation, 2009).

[28] Department of Defense, *Irregular Warfare Joint Operating Concept (JOC)*, (Washington D.C: Department of Defense, 2007).

[29] Ibid.

advanced technology. Success in irregular warfare is centered on winning over the population and establishing enduring partnerships at the local level.[30]

According to the irregular warfare Joint Operating Concept (JOC), there is a range of operations that constitute IW activities. The joint operating concept includes:

> Insurgency, counter insurgency operations (COIN), unconventional warfare, terrorism, counter terrorism, foreign internal defense, stability security, transition and reconstructions (SSTR), strategic communication, psychological operations, civil military operations, information operations, intelligence and counter intelligence (CI) activities, transnational criminal activities (including narcotics trafficking, illicit arms dealing, and illegal financial transactions that support or sustain Irregular Warfare) and Law enforcement activities focused on countering irregular adversaries.[31]

B. PROBLEM STATEMENT

Warfare is undergoing constant evolution due to rapid development of human society. The old ways and means by which to achieve success may be becoming obsolete. The twentieth century has witnessed transformation of warfare from conventional to irregular. There is a noticeable transformation of warfare from the third to the fourth generation.[32] The major conflicts in which the world is embroiled presently are irregular in nature. The Islamic State of Iraq and Syria (ISIS), Taliban, Al-Qaeda, and Maoist Rebels are a few of the major non-state actors playing significant role in fighting modern and well-equipped armies.

These non-state actors are challenging the technological, financial, and organizational superiority of the well-trained armies by using irregular warfare. Irregular warfare is an instrument in the hands of the perpetrators for subduing large armies. The reason for their success is not merely the use of irregular warfare; rather it is ignorance of the regular armies of this emerging trend. Armies around the globe are still more focused and organized to fight conventional war instead of countering irregular warfare. The doctrine, weaponry, and training focus on the third generation of warfare, which is quite different from the contemporary requirement. The nine principles of war are considered

---

[30] Ibid.

[31] Ibid.

[32] William S Lind, "Understanding Fourth Generation Warfare," *Military Review* no. 5 (2004): 12–16.

as guiding principles in conventional warfare for formulation of strategy and execution of battles. In irregular warfare the interpretation of these principles require deliberation and reconsideration. The difference and similarities between the two generations of warfare must be specified for better organizing defense forces. The organizational, training, and executional changes require for defeating irregular warfare cannot be carried out without an in-depth analysis of both types of warfare. The differences and similarities (if any) can help the military decision-maker to reorganize his or her structure of the defense forces.

The other problem in developing an understanding of the transforming nature of warfare is the gap between qualitative and quantitative analysis. Warfare has been studied mostly in separate qualitative and quantitative spheres. There is a void between these two methods of studying warfare that results in misunderstanding at the executional end. The operational research methodology of statistical analysis and combat modeling effectively combined with the qualitative approach can be used for discovering the important factors for winning battles.

C. THESIS ORGANIZATION

1. Research Questions

This research is guided by the following questions.

1. What are battle-winning factors for conventional and irregular warfare?
2. What long-term trends in battles can be identified using historical combat data?
3. What are the major variables that differ between conventional and irregular warfare?
4. What are the means to predict future irregular conflicts based on social development indicators?

2. Methodology

This thesis differs from previous studies on combat in many ways. The focus of this study uses both qualitative and quantitative approaches to find answers to relevant questions on warfare. The study is conducted using statistical analysis of available data

on conventional and irregular warfare. The previous work on the subject has been either qualitatively or quantitatively focused, and mostly in the domain of conventional warfare. Irregular warfare has been neglected by major researchers in the past in favor of a focus on conventional warfare. However, for the last decade, most of the conflicts have been irregular in nature. This thesis examines in depth both types of warfare using both qualitative and quantitative approaches. The focus of this thesis is to transform analytical, mathematical, and combat modeling analysis into simple language so that it can be understood and implemented by executors.

The statistical analysis is done on various datasets separately. First, conventional battles are analyzed based on the CDB90G dataset. The analysis is carried out based on variables selected in accordance with the principles of war to determine factors affecting the outcomes of battle, factors associated with victory, and long-term trends. The U.S. military has not maintained any readily available database for irregular conflicts. Irregular conflicts are evaluated using different open source databases, which include the Armed Conflict Location and Event Data Project (ACLED),[33] Correlates of War Project (COW),[34] and the University of Uppsala conflict database.[35] These datasets are evaluated to determine trends for winning battles, the nature of conflicts, and the likelihood of irregular conflicts in the future.

3. Hypothesis

Conventional and irregular warfare are vastly different; however many military planners seem to have ignored this fact while carrying out planning. The famous Clausewitz maxims, which are applicable in conventional warfare, may be interpreted differently when dealing with irregular warfare. The overlapping of these two different forms of warfare sometime leads to cataclysm. This requires detailed analysis of different battles associated with both forms of warfare to identify the commonalities and

[33] Clionadh Raleigh, Linke Andrew, Hegre Havard and Karlsen Joakim, "Introducing ACLED: An Armed Conflict Location and Event Dataset Special Data Feature," *Journal of Peace Research* 47, no. 5 (2010): 651–660.

[34] Meredith Reid Sarkees, "The COW Typology of War: Defining and Categorizing Wars (Version 4 of the Data)," *Note with Version 4 of the Correlates of War Data* (2010a).

[35] Lotta Harbom, "UCDP Dyadic Dataset Codebook," *Uppsala Conflict Data* (2009).

dissimilarities between the two forms. With identification of these we can determine battle-winning factors.

4. Thesis Organization

The thesis is organized into seven chapters. Chapter II focuses on the literature review and statistical techniques in the field of combat analysis. Chapter III begins with an overview of the principles of war and selection of variables for the purpose of analyzing the conventional warfare datasets. Chapter III incorporates data analysis and descriptive statistics for conventional battles between 1600 and 1982. Chapter IV analyzes three different datasets to determine trends in irregular warfare between 1818 and 2013. Chapter IV deals with statistical analysis of data related to irregular warfare from the Correlates of War dataset. In Chapter VI, we first model the conventional warfare dataset using a classification tree technique to determine important trends and variables in conventional battles. Next, we model an irregular conflict dataset using a boosted tree technique to determine important factors in irregular warfare. Chapter VII provides detailed findings and recommendations for future studies.

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND AND LITERATURE REVIEW

"Study the past if you would define the future"

–Confucius

A. STUDY OF WARFARE

The fundamental reason to study warfare is its continuous and ever-transforming nature and its everlasting impact on human civilization.[36] In Clausewitz's concept, war is a continuity of politics: it certainly leaves an impact on the participants. If an effective prediction had been made available using statistics, the losing side of many battles may have opted to surrender rather than dying. The statistical analysis and theory of combat might provide the means for increasing the odds of victory.

The second fundamental reason for studying war is that it provides an opportunity to analyze complex human conduct in different conditions. There are no simple explanations to these complex dynamics. The study of these conditions can provide insight into war and assist in exploring factors involved in winning the battles. The resources, strategy, and tactics are the common variables normally associated with victory.[37] There are two basic methods used by the analyst to predict warfare: statistical analysis and combat modeling.

1. Statistical Analysis

The study of history should be, as Clausewitz suggested, meant to educate the mind of the future commander, or, more accurately, to guide him in his self-education, not to accompany him to the battlefield; just as a wise teacher guides and stimulates a young man's intellectual development, but is careful not to lead him by the hand for the

---

[36] Azar Gat, *War in Human Civilization* (New York: Oxford University Press, 2006).

[37] C. William Martel, *Victory in War: Foundations of Modern Strategy*. Revised and Expanded Edition (New York: Cambridge University Press, 2011).

rest of his life.[38] The British historian and military analyst Sir Michael Howard observed that studying warfare within a historical perspective helps leaders to gain insight into warfare and make wise decisions.[39]

Statistical analysis is a method widely used for understanding the complexities involved in war and predicting the future based on historical data. Analysts use data based on different battles and campaigns to statistically explore and reveal historical trends. Thus, statistical analysis can give guidance for executing the war, and military leaders should use statistical analysis for their benefit. Analysts in the field have tried to answer inquiries that relate to long-term established principles of war, such as attrition, rates of advance, force ratios, and battle termination rules.[40] According to Helmbold, Roman analyst Vegetius did the first statistical work on warfare in 380 AD to calculate the rate of advance.[41] Helmbold, in his research *"Rates of Advance in Historical Land Combat Operations,"* also compared the works of 34 analysts who studied the rates of advance. All of these analyses focused on providing military leaders information associated with success in battles.

2. Combat Modeling

The second method used by military analysts is combat modeling. A model provides a simplified representation of reality (Figure 1).[42] According to the Defense Modeling and Simulation Office (DMSO), a model is defined as "A physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process."[43] Models can be of different types, such as physical, engineering,

---

38 Peter Paret, Clausewitz in His Time: Essays in the Cultural and Intellectual History of Thinking about War (New York, NY: Berghahn Books, 2014).

39 Michael Howard, "The Causes of Wars," *The Wilson Quarterly* (1984): 90–103.

40 Muzaffer Coban, "Predicting Battle Outcomes with Classification Trees" (master's thesis, Naval Postgraduate School, 2001).

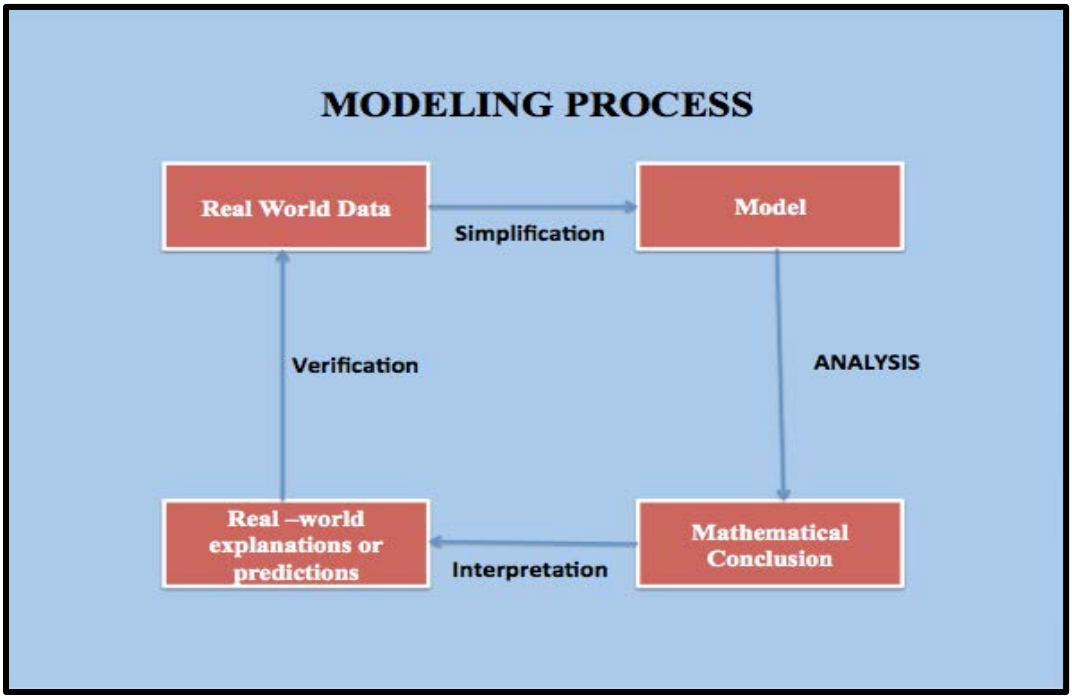
41 Helmbold L Robert, *Rates of Advance in Historical Land Combat Operations.* (CAA-RP-90-1). Bethesda, MD: U.S. Army Concepts Analysis Agency, 1990.

42 James E Coolahan, "A Vision for Modeling and Simulation at APL," *Johns Hopkins APL Technical Digest* 26, no. 4 (2005).

43 Department of Defense, *Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A), Instructions Number 5000.6*, Washington D.C: Department of Defense, 2009.

mathematical, or computer programs depending on the purpose for which they are conceived. Combat models are generalizations of conflicts, their situations, and possible realizations.[44]

Figure 1. Modeling Process Flow



The Kriegsspiel wargames developed in the early nineteenth century by the Prussian Army for purposes of training, planning, and testing are considered the earliest combat models.[45] These models played an important role in the military operations that led to the unification of Germany in 1871. The British engineer Frederick Lanchester formulated a set of mathematical equations during World War I (WWI) to demonstrate opposing force relationships.[46] These equations became the famous Lanchester force-on-force combat model. WWII witnessed the formal development of combat modeling by the Allied forces to achieve war objectives.

---

[44] Milan Vego, "German War Gaming," *Naval War College Review* 65, no. 4 (2012).

[45] Ibid.

[46] Nigel Perry, *Applications of Historical Analyses in Combat Modelling*. Joint Operations Division, Defense Science and Technology Organization (Canberra, Australia: 2011).

However, the revolution of computer technology further expanded this field. Combat models now have become more complicated, and require detailed analysis.[47] According to Macal and North, "many of these models are high-definition, large simulations or agent-based models focusing on modeling conflict by a group of individual entities."[48] These models are useful in defining processes, building systems such as air defense, determining attrition rates, analyzing trade-offs between different factors involved in the acquisition process, determining operational plans, and defining long-term organizational goals.[49] These models can sometimes be misleading, too, since they are just abstractions of reality and not reality in and of themselves. However, it is the analysts who, with human judgment, produce useful results.

B. LITERATURE REVIEW

In the past, various organizations and authors have analyzed the factors for winning battles based on historical data and combat models. Some previous studies of combat using historical data are discussed in the succeeding paragraphs.

1. Historical Evaluation and Research Organization (HERO)

From 1983 to 1984, Historical Evaluation and Research Organization (HERO) prepared a database of 601 battles and engagements from 1600 to 1973 AD for the U.S. Army Concepts Analysis Agency (CAA).[50] The HERO database consists of seven tables covering battle identification, operational and environmental variables, strengths and losses, intangible factors, outcome, factors affecting the outcome, combat forms, and scheme of resolution.[51]

---

[47] Andreas Tolk, *Engineering Principles of Combat Modeling and Distributed* Simulation (Hoboken, NJ: John Wiley & Sons, 2012).

[48] Charles M Macal, and North J Michael, "Tutorial on Agent-based Modelling and Simulation," *Journal of Simulation* 4, no. 3 (2010): 151–162.

[49]Andreas Tolk, *Engineering Principles of Combat Modeling and Distributed* Simulation (Hoboken, NJ: John Wiley & Sons, 2012)..

[50] Trevor Nevitt Dupuy, Numbers, *predictions, and war: Using History to Evaluate Combat Factors and Predict the Outcome of Battles* (Virginia: NOVA Publications 1985).

[51] Helmbold L Robert. and Khan A Aqeel, *Combat History Analysis Study Effort (CHASE)*. (CAA-TP-86-2) (Bethesda, MD: U.S. Army Concepts Analysis Agency, 1986).

There is an immense amount of data captured on battle experience in this database. The time spans a broad range of technologies and hence should allow analysts to discover important trends. The database is mainly representative of short, pitched-land combat battles fought by division and corps-sized military formations during the nineteenth and twentieth centuries in Europe and North America. It does not contain any sea or air battles; sieges of heavily fortified positions; action from the Korean, Malayan, Algerian, or Vietnamese Wars; or extensive data on WWII. It contains hardly any Asian, African, Middle East, or South American wars, except for the Arab-Israeli wars of 1967–1973. The following table summarizes the HERO database:

Table 1. Hero Database Descriptive Statistics[52]

SCOPE OF DATABASE	
Total Number of Battles	601
Battle Dates	1600–1973 A.D.
Total Engaged Strength	89 $x10^6$ troops
Total Engaged Troop Days	1.1 $x10^9$ troops
Total Battle Casualties	19 x 10^6 troops
Average Casualty Rate	2% per troop day
Total Battle Days	2,300 days
Total Area Gained by Attacker	1.3 $x10^6$ sq km

2. Combat History Analysis Study Effort (CHASE)

The database prepared by HERO was unique and of great potential, yet CAA was unable to use it directly for studies and analysis because it did not provide quantitative trends and interrelations.[53] As a result, CAA established the CHASE project in 1984,[54] with the objective of searching for historically-based quantitative results for use in military operations research, concept formulation, war-gaming, studies, and analyses.

[52] Ibid.
[53] Ibid.
[54] Ibid.

These five essential elements of analysis guided the study: factors associated with victory in battle, long-term trends in historical combat data, factors influencing rate of advance, the effect of air support on battle outcome, and preparation of databases for quantitative analysis.[55]

CHASE carried out a descriptive analysis of the aforementioned factors associated with select battles, measuring battle outcome based on the six "surrogate" variables. A detail description of these variables is discussed in succeeding paragraphs:

(1) Force ratio (FR) is defined as the ratio of the attacker's number of personnel (X_0) to the defender's personnel (Y_0).

$FR = X_0 / Y_0$

Bitterness (EPS) is derived on Lanchester Equations. It is defined as

EPS = SQRT (Attacker's Fractional Loss (FX)* Defender's Fractional Loss (FY))
Attacker's Fractional Loss (FX) is defined as
Attacker Number of Casualties / Attacker Number of Personnel
Defender's Fractional Loss (FY) is defined as
FY = Defender's Number of Casualties / Defender Number of Personnel

(2) Casualty exchange ratio (CER) is defined as the ratio of the attacker's casualties (CX) to the defender's casualties (CY).

CER = CX/ CY

(3) Fractional exchange ratio (FER) is defined as the ratio of the attacker's to the defender's fractional losses.

FER=FX/FY

(4) Advantage (ADV) is defender empirical advantage parameter.
ADV=(1/2)*LOG (FER)

(5) Residual advantage (RESADV) is residual value of advantage (ADV) after the average effects of any differences in force ratio (FR) values are removed.[56] It is written as

RESADV= ADV $- a - b$ *LOG (FR)

[55] Ibid.
[56] Ibid.

a and *b* are the regression coefficients.

These variables, since the completion of the CHASE study, have been used by various authors to study the outcome of war. This thesis also considers these variables.

a. Non–WWII Battles Data Subset

The CHASE study used logistic regression as the principal technique to assess the degree to which the surrogate variables are related to conventional battle outcome categories (win, lose, or draw). The basic results from the logistic regression computation are presented in Table 2. The column labeled L(0) gives the log likelihood value from logistic regression. The columns labeled a (1,0), a (1,1), a (2,0), and a (2,1) give the maximum likelihood parameter values of the logistic function of the non-WWII data subset used. The columns labeled SD (1,0), SD (1,1), SD (2,0), and SD (2,1) give the standard deviations of the maximum likelihood parameters.

Table 2. CHASE Logistic Regression Results[57]

LOGISTIC REGRESSION RESULTS											
Independent Variable	Number of data points	L(0)	Max.L	a (1,0)	SD (1,0)	a (1,1)	SD (1,1)	a (2,0)	SD (2,0)	a (2,1)	SD (2,1)
ADV	427	-469	-219	-1.527	0.26	-3.783	0.80	0.247	0.15	-5.997	0.63
LOG(FER)	427	-469	-219	-1.522	0.26	-1.733	0.37	0.242	0.15	-2.770	0.29
RESADV	427	-469	-222	-1.214	0.24	-3.477	0.78	0.770	0.16	-6.136	0.63
LOG(CER)	427	-469	-239	-1.248	0.26	-1.225	0.32	0.888	0.16	-2.308	0.24
LOG(EPS)	427	-469	-354	-1.832	0.54	0.013	0.22	0.905	0.27	0.164	0.11
LOG(FR)	435	-478	-362	-1.892	0.25	0.364	0.30	0.468	0.11	0.326	0.16

[57] Ibid.

b. ***Ranking of Variables***

The increase in log likelihood expressed through the quantity (Max.L - L(0)) provides a rough measure of the relative quality of the logistic regression fit. For this measure, the variables ADV, LOG (FER), and RESADV are approximately tied for best fit. The variable LOG (CER) is next best. The variables LOG (EPS) and LOG (FR) are approximately tied for worst fit. Table 2 also shows that the variables ADV and LOG (FER) are essentially equivalent with regard to logistic regression. Surprisingly, force ratio is an inadequate predictor of victory in battle. Both advantage and fractional exchange ratios are much more closely related to victory than is the force ratio.

3. McQuie's Benchmarks Study

McQuie conducted a study in 1988 to identify quantitative measures for comparing historical data and the results of simulations and war games.[58] The study was carried out on a dataset comprising 601 battles generated from various studies over the past 25 years by DuPuy.[59] These data have been assembled under a contract with CAA and now constitute the Army's database of historical battles. McQuie argued future battles would not replicate past ones; however, predicting factors in winning future battles requires credible comparisons with past battles. McQuie computed 28 different ratios and rates from the data and gave measures of variability, as shown in Table 3. The historical dataset was not complete, so historians estimated the missing information.

---

[58] Robert McQuine, *Historical Characteristics of Combat for Wargames*. Bethesda, MD: U.S. Army Concepts Analysis Agency, 1988.

[59] Dupuy N. Trevor, The Dupuy Institute2015. http://www.dupuyinstitute.org/tndupuy.htm.

Table 3. Measures of Variability[60]

MEASURES OF VARIABILITY					
Characteristic	Type of Characteristic	Year	Attacker & Defender	Value of Characteristic	Ratio of High to Low Values
Force Ratio Men (Atkr:Dfdr)	High:	1967	Egypt: Israel	17:1	57:1
	Low:	1945	Japan: USA	0.3:1	
Force Ratio Artillery (Atkr :Dfdr)	High:	1945	USA: Japan	50:1	450:1
	Low:	1948	Israel: Syria	0.11:1	
Mortar Density Dfdr (wpns/km)	High:	1943	Britain: Germany	132	730:1
	Low:	1973	Egypt: Israel	0.19	
Artillery Density Atkr (wpns/km)	High:	1944	USA: Japan	444	2200:1
	Low:	1948	Israel: Jordan	0.2	
Casualty Rate Atkr (%per day)	High:	1945	USA: Japan	96%	740:1
	Low:	1944	Britain: Germany	0.13%	
Tank Loss Rate Atkr (%per day)	High:	1967	Israel: Syria	92%	150:1
	Low:	1944	USA: Germany	0.63%	
Advance Rate (km per day)	High:	1967	Israel: Egypt	45	450:1
	Low:	1945	USA: Japan	0.1	

[60] Robert McQuine, *Historical Characteristics of Combat for Wargames*. Bethesda, MD: U.S. Army Concepts Analysis Agency, 1988.

4. Bracken's Attrition Model of Ardennes Campaign

In 1995, Bracken analyzed the Ardennes campaign carried out during World War II.[61] Bracken presented four Lanchester models of the campaign and estimated their parameters.[62] Bracken formed homogenous weighted models that comprised elements such as tanks, armored personnel carriers, artillery, and manpower to yield a measure of strength to the Allied and German forces. Two of Bracken's models consisted of five parameters: Allied individual effectiveness, German individual effectiveness, exponent of shooting force, exponent of target force, and a tactical parameter reflecting which side was defending or attacking. The other two Bracken models were formed after removing the "tactical parameter, which is not generally known prior to a battle." Bracken concluded that the "Lanchester linear model fits the Ardennes campaign data." For the model consisting only of combat forces, Allied individual effectiveness was greater than the Germans'. For the model consisting of both combat and support forces, Allied and German effectiveness was equal; the Allies used more forces in combat support, which enhanced their individual combat efficiency.

5. Clemens's Study of the Battle of Kursk

Clemens (1997) examined the validity of the Lanchester Models applied to modern warfare.[63] Clemens applied the Lanchester models to the Battle of Kursk data.[64] Clemens uses two estimation techniques, linear regression and Newton–Raphson iteration. His analysis also explores the presented model in matrix form and compares the matrix solution to the scalar solution. In his study, he concludes that neither Lanchester linear nor square models fit the data, whereas the Lanchester logarithmic model in both scalar and matrix forms fits.

---

61 Jerome Bracken, "Lanchester Models of the Ardennes Campaign," *Naval Research Logistic* 42, no. 4 (1995): 559–577.

62 Ibid.

63 SC Clemens, "The Application of Lanchester Models to the Battle of Kursk," *Unpublished Manuscript, Yale University* 5, (1997).

64 Ibid.

6. Fricker's Study of the Ardennes Campaign

In 1998, Fricker reexamined the Ardennes campaign and Bracken's model using the Lanchester equations.[65] He used a different approach than Bracken. Fricker extended Bracken's analysis by using linear regression on the dataset for the entire campaign. Fricker explored the complete battle dataset in contrast to Bracken, who only evaluated the first ten days of battle data.[66] Fricker also included data on air sorties in his study and evaluated the impact of air on the battle. Fricker's conclusions are different from Bracken's: he concluded that Lanchester's basic models (linear and square) do not fit the Ardennes campaign, which is in direct contrast to the Bracken study.

7. Hartley and Helmbold's Application of the Lanchester Square Model

Hartley and Helmbold utilized linear regression to test whether the Lanchester square model applies to the Inchon-Seoul campaign of the Korean War.[67] Their dataset consisted of manpower only, and they attempted to model just United States casualties. Hartley and Helmbold use three analysis techniques to examine the data: linear regression, the Akaike Information Criterion (AIC), and Bozdogan's consistent AIC (CAIC).[68] In addition, they introduced the use of change points at certain phases in the campaign and then refitted the model at each of these change points. They concluded that the data do not fit a Lanchester square law with a constant coefficient and that more real data are required for validating the Lanchester square law.[69]

---

65 Ronald D Fricker, "Attrition Models of the Ardennes Campaign," *Naval Research Logistics* 45, no. 1 (1998): 1–22.

66 Ibid.

67 Dean S. Hartley and Robert L. Helmbold, "Validating Lanchester's Square Law and Other Attrition Models," *Naval Research Logistics (NRL)* 42, no. 4 (1995), 609–633.

68 Ibid.

69 Ibid.

8. Lucas and Turkes's Study on the Battles of Kursk and Ardennes

Lucas and Turkes extend previous research by validating Lanchester's equation with real data.[70] They developed an understanding of highly aggregated attrition by fitting the homogenous generalized Lanchester model to the time-phased data of the battles of Kursk and Ardennes. Their study assisted in developing the campaign-level simulation. They followed a new method to discover the optimal factor values and to develop an insight into how various factors combined explain the battle. Their study concluded that most of the Lanchester model fit the data well if the battle is broken into phases. The study also found that choosing correct coefficients would yield consistent results irrespective of the law used. The study also proved that the constant attrition coefficient does not fit the Lanchester law very well.[71]

9. Lucas and Dinges' Study on the Battle of Kursk

Lucas and Dinges study the effect of battle circumstances on fitting Lanchester equations to the battle of Kursk. Lucas and Dinges used the Center for Army Analysis's data for the southern front of the battle of Kursk. This dataset is unique since it captures the daily combat status of all divisional-level units. The dataset categorized units into three different categories: first, all combat forces in the campaign; Second, all combat forces within contact; and third, all combat forces that are actively engaged. They concluded that more of the variation in casualties during the battle of Kursk is explained by the category of the force considered and the phases of the battle than by the Lanchester variant used. Better fits were obtained by using forces that were actively engaged. Moreover, they were able to improve results of their model by dividing battles according to natural phases. Finally, they concluded that Lanchester's linear law fits better than the logarithmic law.[72]

---

[70] Thomas W. Lucas and Turkes Turker, "Fitting Lanchester Equations to the Battles of Kursk and Ardennes," *Naval Research Logistics (NRL)* 51, no. 1 (2004), 95–116.

[71] Ibid.

[72] Thomas W. Lucas and John A. Dinges, "The Effect of Battle Circumstances on Fitting Lanchester Equations to the Battle of Kursk," *Military Operations Research* 9, no. 2 (2004), 17–30.

10. Yigit's Thesis

Yigit explored "important factors in battle through a statistical analysis of data from major historical battles."[73] Yigit carried out a comprehensive analysis seeking patterns, trends, and relations in combat. Yigit organized the dataset by chronological order of the campaigns and focused on exploring force ratio, dispersion, and daily casualty rates. The famous force ratio of 3:1 has always dominated military planning and remains a point of discussion as one of the most important variables in warfare. Yigit explored this important factor and concluded that in 68%of the battles, a force ratio of 3:1 led to victory. His findings on force ratio are tabulated below:

Table 4. Probability of Attacker Win Given Force Ratio[74]

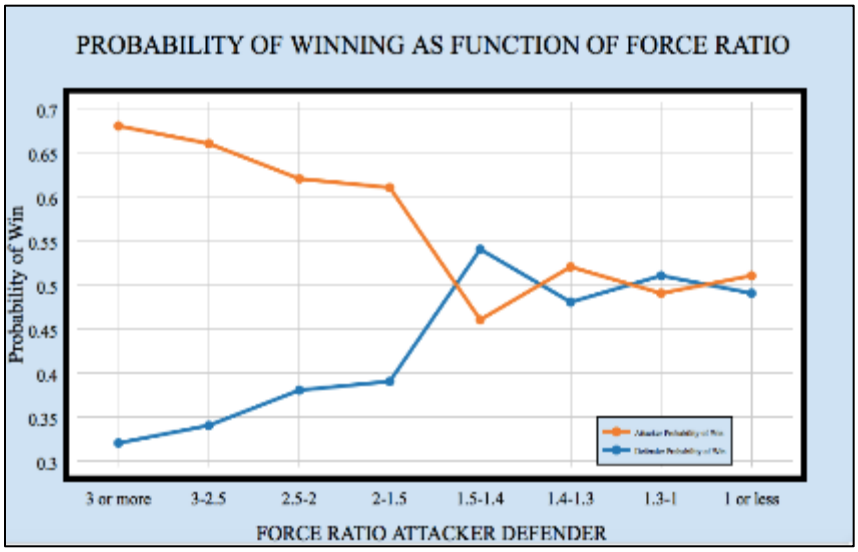
PROBABILITY OF ATTACKER WIN							
3 or More	3-2.5	2.5-2	2-1.5	1.5-1.4	1.4-1.3	1.3-1-1	1 or less
0.679	0.657	0.619	0.613	0.461	0.523	0.485	0.513

From Yigit's findings, if an attacker has numerical superiority, success is likely. Yigit concluded with the probability of winning based on force ratio in his statistical analysis, which is graphically represented in Figure 2.

[73] Faruk Yigit, "Finding the Important Factors in Battle Outcomes: A Statistical Exploration of Data from Major Battles" (master's thesis, Naval Postgraduate School, 2000).

[74] Ibid.

Figure 2. Probability of Winning[75]



While analyzing dispersion on the battlefield, Yigit observed that dispersion increased as casualties decreased, despite increasing lethality of weapons.

Table 5. Density and Dispersion of Troops through History (Man per Km^2)[76]

DENSITY AND DISPERSION					
Year	Battle	Density of the attacker	Density of the defender	Dispersion of the attacker	Dispersion of the Defender
1805	Napoleonic War	4494.420	3232.00	222.498	309.406
1861	American Civil War	3378.293	2838.674	296.007	352.77
1870	Franco –Russian War	2125.886	2838.674	296.007	352.77
1914	WW-I	345.168	186.745	28975.142	5354.895
1944	WW-II	72.584	30.902	13777.154	32360.108
1973	Arab –Israel 1973	35.031	29.928	28546.274	33413.389

[75] Ibid.
[76] Ibid.

Yigit identified three trends in daily casualty rates that have generally declined over the past four centuries and almost leveled off at the rates experienced in WWII and the Arab-Israel wars. The casualty rates of the attackers were almost always lower than those of the defenders. Yigit also observed that casualty values decreased as the unit size in the battle increased. The daily casualty rates of small units are higher than those of the large forces under the same circumstances. He identified two principal reasons: first, small combat forces had very few individuals who were not related to combat, and second, control over the units is enhanced in the battlefield when the size of the unit increased.

11. Coban's Thesis

Coban used classification trees and divided the data into response, objective, and relative variables to predict outcomes of battles.[77] Coban divided the dataset into five subsets, as shown in Table 6.

Table 6. Division of Data into Subsets[78]

DIVISION OF DATA INTO SUBSETS						
SUBSET NO	SUBSET	SIZE	TRAINING SET	SIZE	TEST SET	SIZE
1	1600-1847	164	1600-1799	109	1799-1847	55
2	1805-1918	260	1805-1915	178	1916-1918	82
3	1920-1945	202	1920-1944	131	1944-1945	71
4	1940-1982	223	1940-1948	150	1950-1982	73
5	1600-1982	658	1600-1944	435	1940-1982	223

[77] Muzaffer Coban, "Predicting Battle Outcomes with Classification Trees" (master's thesis, Naval Postgraduate School, 2001).

[78] Ibid.

Coban concluded that models using only objective variables resulted in high misclassification rates, whereas models based on both objective and relative variables resulted in low misclassification rates. He concluded that objective variables alone are not reliable predictors of battle outcome. Furthermore, the variables affecting outcomes have changed through history. Specifically, earlier battles have fewer relevancies with modern era battles.

THIS PAGE INTENTIONALLY LEFT BLANK

III. STATISTICAL EXPLORATION

A. DATA ANALYSIS FOR WARFARE STUDIES

Military historians predict future warfare based on the prism of history using analytical and qualitative analysis. During the nineteenth century, with the evolution of modern mathematics, statistics, and operations research techniques, warfare analysis became more quantitative. Because history is written by the victors, data are often biased. The fog of war, too, makes accurate data collection difficult. However, this limitation can be mitigated by applying British Military historian Sir Michael Howard's concept of studying both the depth and breadth of warfare.[79] Statistically speaking, we can mitigate these effects by sampling a large amount of data (depth) and evaluating a significant number of variables for analysis (breadth). This may reduce some forms of bias and will hopefully give a fairly balanced picture.

Correlating battle data can help analysts predict future trends in warfare. This research endeavors to determine the important variables in winning battles and war for both conventional and irregular warfare. Different datasets are explored during this study: the first dataset, CDB90G, is of conventional battles, the second dataset, from ACLED, is of politically violent conflicts, and the third dataset, from the Correlate of War (COW) project, is of both conventional and irregular warfare.

B. BRIEF OVERVIEW OF CDB90G DATASET

For the analysis of conventional battles, the CDB90G dataset is analyzed. This dataset is not randomly selected and the compiler of data does not provide information on selection criteria for battles to be included in the dataset. This dataset includes 664 battles between the seventeenth and twentieth centuries. The first battle in the dataset is the battle of Nieuport, which was carried out during the Netherlands war of independence at Spanish Flanders. The last battle in the dataset is the 1982 battle of Lebanon, which took place at Bekka. The dataset has a total of 206 variables; however, analysis does not

[79] Howard, Michael. "The Causes of Wars," *The Wilson Quarterly* (1984): 90–103.

include all of them during modeling. The variables are selected based on theories of war and professional military knowledge. Before the selection of each variable, it is essential to review the dataset in order to gain insight into missing values. There are a total of 46,456 missing values in the dataset, details of which are given in Table 7. These missing values appear in order of variable name. A codebook for dataset CDB90G contains variable names and relevant details. The missing values can cause biases in analysis. However, tree modeling can effectively mitigate missing values.

Table 7. Missing Values in Data

MISSING VALUES IN DATA						
NationA	NationD	ATPBYR2	ATPBMN2	ATPBDA2	ATPBHR2	ATPEYR2
4	4	564	564	564	564	564
ATPEMN2	ATPEDA2	ATPEHR2	ATPBYR3	ATPBMN3	ATPBDA3	ATPBHR3
564	564	564	623	623	623	623
ATPEYR3	ATPEMN3	ATPEDA3	ATPEHR3	ATPBYR4	ATPBMN4	ATPBDA4
623	623	623	623	650	650	650
ATPBHR4	ATPEYR4	ATPEMN4	ATPEDA4	ATPEHR4	ATPBYR5	ATPBMN5
650	650	650	650	650	656	656
ATPBDA5	ATPBHR5	ATPEYR5	ATPEMN5	ATPEDA5	ATPEHR5	ATPBYR6
656	656	656	656	656	656	659
ATPBMN6	ATPBDA6	ATPBHR6	ATPEYR6	ATPEMN6	ATPEDA6	ATPEHR6
659	659	659	659	659	659	659
ATPBYR7	ATPBMN7	ATPBDA7	ATPBHR7	ATPEYR7	ATPEMN7	ATPEDA7
663	663	663	663	663	663	663
ATPEHR7	ATPBYR8	ATPBMN8	ATPBDA8	ATPBHR8	ATPEYR8	ATPEMN8
663	663	663	663	663	663	663
ATPEDA8	ATPEHR8	ATPBYR9	ATPBMN9	ATPBDA9	ATPBHR9	ATPEYR9
663	663	664	664	664	664	664
ATPEMN9	ATPEDA9	ATPEHR9	ATPBYR10	ATPBMN10	ATPBDA10	ATPBHR10
664	664	664	664	664	664	664
ATPEYR10	ATPEMN10	ATPEDA10	ATPEHR10			
664	664	664	664			

In addition to missing values, there are 101 variables whose results for 24,210 values are unknown. There is a difference between missing and unknown values. While there is no numeric descriptor for missing values, unknown values for different variables have numeric descriptors. Most of the unknown and missing values are not required for our analysis. However, it is essential for future studies that these be identified and documented. Details of unknown values are given in Table 8, and the corresponding codebook contains detail regarding the variables.

Table 8. Unknown Values in Data

UNKNOWN VALUES IN DATA						
COA	WOFA1	WOFD1	YR1	MO1	DA1	Hr1
42	17	47	629	629	629	652
WOFA2	WOFD2	YR2	MO2	DA2	HR2	WOFA3
622	622	629	629	629	653	660
WOFD3	YR3	MO3	DA3	HR3	Front	Depth
660	660	660	660	663	50	50
Time	AERO	STRA	INTSTA	REPRA	CASA	FINSTA
50	38	-1	128	272	6	188
STRD	INTSTD	RERPD	CASD	FINSTD	CAVA	TANKA
1	135	275	4	181	81	19
LTA	MBTA	ARTYA	FLYA	CTANKA	CARTYA	CFLYA
94	94	107	132	111	434	326
CAVD	TANKD	LTD	MBTD	ARTYD	FLYD	CTANKD
79	20	91	91	121	144	101
CARTYD	CFLYD	CEA	LEADA	TRNGA	MORALA	LOGSA
406	269	38	38	38	38	38
MOMNTA	INTELA	TECHA	INITA	WINA	KMDA	ACHA
38	38	38	38	2	26	38
ACHD	CRIT	QUALA	RESA	MOBILA	AIRA	FPREPA
38	62	38	38	38	38	38
WXA	TERRA	LEADAA	PLANA	SURPA	MANA	LOGSAA
38	38	38	38	38	38	38
FORTSA	DEEPA	PR1A1	PRIA2	PR1A3	SECA1	SECA2
38	38	11	479	587	502	577
SECA3	RESOA1	RESOA2	RESOA3	PRID1	PRID2	PRID3
660	24	277	588	40	424	634
SECD1	SECD2	SECD3	RESOD1	RESOD2	RESOD3	ATPHR1
652	658	664	118	532	651	319
ATPEHRI	ATPBHR2	ATPEHR2				
320	1	1				

C. PRINCIPLES OF WAR

Theorists of military science have pursued through history primary laws or theories that explain both the relations of military forces in combat and the outcomes of battles.[80] Jomini, Clausewitz, Mahan, and Fuller are prominent military thinkers who carried out qualitative analysis on historical battles. On the other hand, theorists such as Lanchester carried out quantitative analysis of warfare. These two approaches for study are distinct, but the common foundation on which they stand aims to determine conditions that facilitate attaining victory.

Military literature is full of theoretical observations on warfare. Great captains of warfare such as Napoleon have always benefitted from the theories of their predecessors to ensure success in battle. Napoleon found inspiration from the writings of *Reveries on the Art of War* by Maurice de Saxe and *Instructions to His General* by Frederick the Great of Prussia. Napoleon, although unable to produce a scriptural form of war theory, established concepts on the battlefield in an articulated manner. In his correspondence and recorded statements, he made it clear that his concepts of war had been derived from studying the campaigns of earlier great generals.[81] Henry Jomini, a Swedish officer in Napoleon's army, endeavored to explain Napoleon's theories and tried to model them.

Carl Von Clausewitz, a contemporary of Jomini's, wrote about many aspects of war, but his approach focused on two particular aspects, activities and characteristics of war; some of these characteristics included level of violence and passion, human behavior, and politics. Clausewitz's concept centers on a ratio between two forces based on number of troops, variable circumstances, and quality of force. The greater the number favors the red and at one the forces are equal.

Contemporary theorists developed models based on historical records, which have helped generations of military students understand warfare from a historical perspective and for winning battles. The first great American theorist to compile his own rules

---

80 Earle Mead Edward, Craig Alexander Gordon, and Gilbert Felix, *Makers of Modern Strategy: Military thought from Machiavelli to Hitler* (Princeton, NJ: University Press, 1943).

81 GC. D, Aguilar, Napoleon's Maxims of War with notes by General Burnod (*Pennsylvania*, Pa: David McKay, 1902).

relevant to military theory was Dennis Hart Mahan. Although he never tried to produce a theory, he is highly respected for his work. Another military theorist who inspired generations through his work is Helmuth von Moltke. He gave concept to a combined tactical defense with strategic offensive. John F. C. Fuller, the greatest military thinker of the twentieth century, synthesized the concepts of earlier military theorists. Fuller, as a British officer, studied military history, particularly the campaigns of Napoleon, seeking insight into the basic principles of war.[82] In 1921, Fuller wrote *The Principles of War*, proposing the following fundamentals of battle.[83]

$$Outcome = \frac{(Nr * Vr * Qr)}{(Nb * Vb * Qb)}$$

where

b = indices of blue force,

r = indices of red force,

N= number of troops,

V = variable circumstances,

Q= quality of force.

1. Objective

The objective can also be described as the *mission*, *aim*, or *purpose*. Objective is unquestionably the most important of all the principles of war.[84] Without an objective, all other principles of war are invalid. The objective of war should be clear at all echelons: policy makers, public, and the armed forces. Once the objective has been stated and understood, the problem is simplified.

---

[82] John Frederick Charles Fuller, *The Conduct of War, 1789–1961: A Study of the Impact of the French, Industrial, and Russian Revolutions on War and its Conduct* (Boston, MA: Da Capo Press, 1992).

[83] Ibid.

[84] Ibid.

2. Mass

Mass measures force superiority at the point of contact. In other words, it is superiority that must be sustained as long as the situation requires. Mass is not only numerical superiority, but also a combination of firepower, skills, weapons, discipline, resolve, leadership, administration, and morale.[85]

3. Offensive

The offensive is simply taking the war to the enemy. The ultimate goal for waging war is to attain victory, which can only be achieved through offensive spirit. Well-organized army formations, well-equipped air forces, and large naval fleets are useless if no offensive spirit exists. Victory is not won by passive defense, which merely averts defeat.[86]

4. Surprise

Surprise is considered the greatest weapon in war. It is essentially one of the most effective methods for obtaining victory. Surprise is best defined as the creation of unexpected situations for which the enemy has not properly prepared.[87] Factors of surprise include secrecy, preparation, rapidity of execution, and deception.

5. Security

The prevention of hostile interference is known as security. The mission of security is to prevent plans from falling into enemy hands and, thus, provide freedom of movement. Security means not only denial of information to the enemy but also the ability to obtain information about the enemy.[88]

---

[85] Ibid.
[86] Ibid.
[87] Ibid.
[88] Ibid.

6. Movement

Movement is the ability to change the location of forces. Many successful commanders of the past were able to defeat their enemies due to their ability to rapidly move forces. Movement initiates mass. It is the way in which a force reaches an area to achieve an objective. If a commander is able to concentrate a force at a threatened point, he will have more opportunities to achieve victory. Movement is dependent on logistics and flexibility of planning.[89]

7. Economy of Force

The complement to mass is economy of force, which means employing the fewest resources required at a decisive time and place. Economy of force does not denote stinginess but intelligence to skillfully use a force. Another term for economy of force is *force optimization*. Its application involves questioning what is and what is not necessary. This principle deals with the two demands of dispersion and concentration.

8. Cooperation

Cooperation is a unifying principle of war. Whereas the objective designates a common aim, cooperation denotes a common methodology for achieving an objective. It is a fusion of forces to produce a desired result through a common understanding. Cooperation brings the full power of the nation against the enemy's military, political, economic, financial, and psychological domains.[90]

9. Simplicity

Simplicity allows observance of all other principles of war. A simple plan is easier to execute than a complex one. Simple and direct methods usually preserve the elasticity needed to meet the ever-changing situations of conflicts.[91]

---

89 Ibid.
90 Ibid.
91 Ibid.

D. PRINCIPLES OF WAR: THE U.S. MILITARY

Fuller's nine principles of war laid the foundation for theoretical studies of warfare in the twentieth century. These principles are interesting guidelines for battles, but are neither sacred nor immutable. Blindly following these principles may not always lead to victory because war is stochastic in nature and is affected by numerous factors. However, based on these principles, different armies have developed principles of war that are applicable to their own situations—ones tailored to national needs.

Warfare has evolved over time; social events and the development of society have determined the evolution of conflicts. The conflicts that most armies face today are entirely different from conflicts of the past. The principles that Fuller illustrated were primarily for conventional conflicts in which adversaries were nation states. Present conflicts are more irregular in nature—whether in Afghanistan, Iraq, Syria, or Africa—and more ideologically based than past conflicts. The basic principles of war may not be applicable in this kind of warfare. This leads to debate among military historians, theorists, and researchers over the value of qualitative and quantitative means of analysis. However, this thesis uses both qualitative and quantitative methods to assess the applicability of these principles in the present environment, based on available data. The U.S. Armed Forces have also increased the number of principles of war from nine to twelve. The three new principles of war, which have been added to the nine old principles, are restraint, perseverance, and legitimacy.

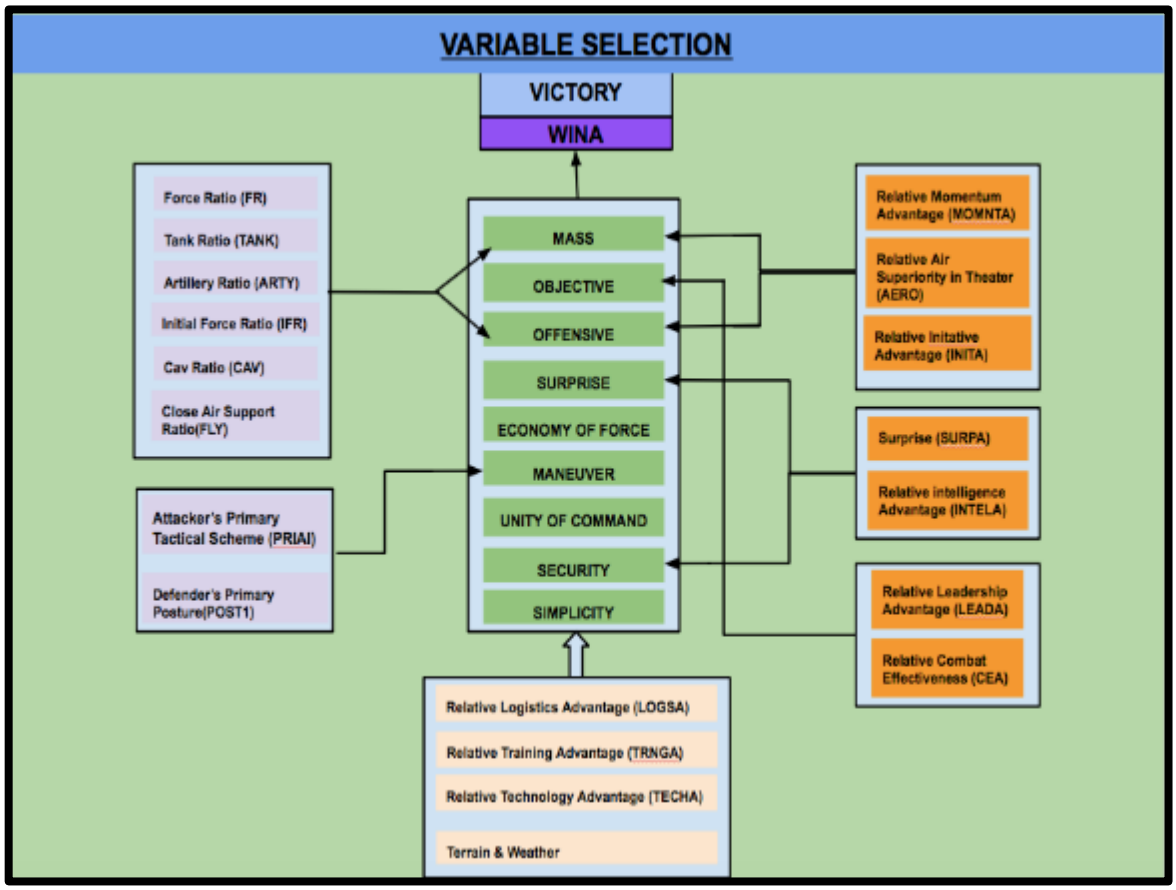
E. SELECTION OF VARIABLES

The Center for Army Analysis (CAA) CDB90G dataset is very comprehensive, with 206 variables, some of which are not relevant for our analysis purpose. To determine favorable conditions for victory, it is necessary that relevant variables should be selected. This selection of variables is debatable because different analysts may have different criteria for judging. In order to overcome individual biases and preferences, the variables selected for modeling purpose have been intrinsically drawn based on the principles of war. This cautious approach has been taken based on an objective of study that is to bridge the gap between qualitative and quantitative approaches to warfare. These

principles were initially founded by Fuller based on his qualitative analysis of warfare. Most of the authors who have written on these principles also adopted a qualitative approach. However, modern operations research techniques can be effectively utilized to assess the efficacy of these principles.

The ultimate aim for war is imposing one's will on the enemy, which can only be achieved through victory. In the dataset, victory is represented by a response variable denoted by "WINA." The response variable is predicted based on objective and relative variables along with two types of conditions, weather and temperature. Three types of models are constructed based on objective variables, relative variables, and conditions. The first type of model is formed based only on objective variables. The second type of model includes both objective and relative variables. The third type of model includes the variables for the first and second type of model along with additional relative variables and a few environmental conditions. In Figure 3, the relationship between different types of variables and the principles of war is graphically illustrated. The left side includes objective variables used for Model 1. The right side includes relative variables, and on the bottom there are a few more relative variables with the two additional conditions.

Figure 3. Variable Selection Based on Principles of War



F. DESCRIPTIVE STATISTICS

Descriptive statistics is the expression given to the examination of data that helps explain, display, or summarize it in a simpler, more meaningful way.[92] By using the descriptive statistics techniques such as box plots, histograms, and geographical graphs, data can be visualized and summarized in meaningful ways. In the succeeding paragraphs, descriptive statistics techniques are used to gain insight into the data.

1. Data Augmentation

The CDB90G dataset from the Center for Army Analysis (CAA) contains descriptive locations of battles; however, the precise longitude and latitude along with the

[92] Thomas H. Wonnacott Wonnacott H Thomas, and Wonnacott J Ronald, *Introductory Statistics*. Vol. 19690 (New York, NY: John Wiley& Sons, 1972).

year in which battles took place were sometimes missing. What the dataset listed were regions, countries, cities, or general locations. We augmented the dataset with specific longitude and latitude as well as year for each battle. The numbers of battles by commencement year are listed in Table 9.

Table 9. Number of Battles in Different Centuries

NUMBER OF BATTLES IN CHRONOLOGICAL ORDER			
YEAR	Number of Battles	First Battle of Century	Last Battle of Century
1600-1697	48	1600	1697
1700-1799	65	1704	1799
1800-1899	126	1800	1899
1900-1982	425	1900	1982
Total	664		

The largest number of battles, as shown in Table 9, come from the twentieth century. Another noticeable trend is that the number of battles increases with every new century. There are two reasons for this. First, with the evolution of the human race came an increasing struggle for control of resources, which reached its zenith in the twentieth century. Second, advancements in archiving increase the number of battles recorded. For instance, the battle accounts for the twentieth century are generally documented in more detail than those of the sixteenth century. Based on this argument, it is possible that there may have been more battles fought during the sixteenth century than the dataset reflects. Another insight, which we can obtain from the dataset, is identifying the years in which the most battles took place. Figure 4 represents the intensity of battles over time in the CDB90G dataset.

Figure 4. Battle Intensity over Time

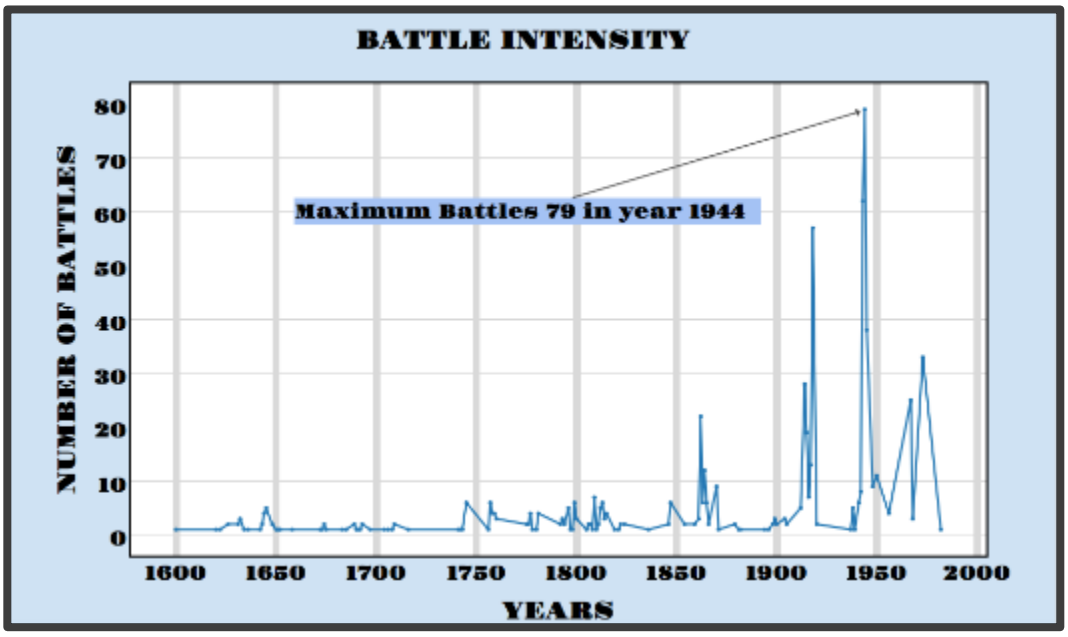
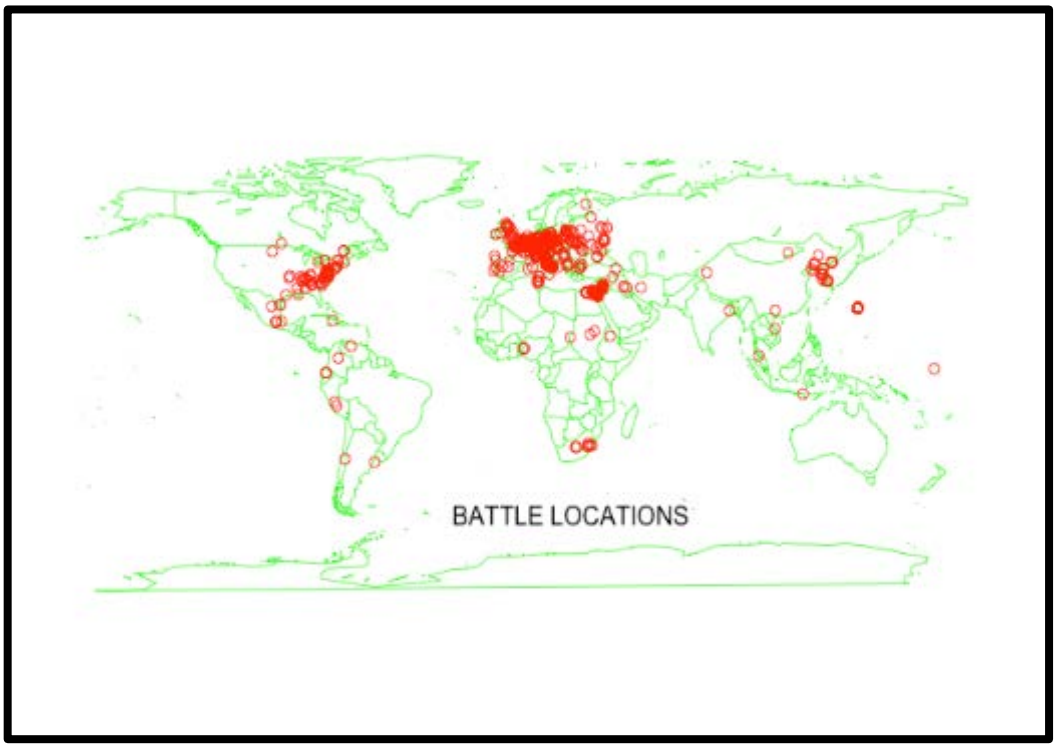


Figure 4 indicates most battles were fought between 1900 and 1950. This is the time frame that reflects two world wars. Although the graph seems quite intuitive and logical, there are certain battles for which records have not been available due to difficulty in data collection. Viewing the geographic data on a map reveals areas which may be inadequately represented in the dataset.

Figure 5. Battle Locations Based on CDB90G Dataset



The data reveal that most of the battles in the dataset were carried out in Europe. The dataset reflects 418 battles in Europe, 90 in the Middle East, 82 in the United States, 57 in the Pacific, eight in Africa, seven in South America, one in India, and one in China. This does not mean that the rest of the world was peaceful; there were wars in different regions, but they may not have been documented well or were of less interest to the U.S. Army.

Figure 6. Battle Locations With Geographical Distributions

2. Battles per Campaign

The campaigns of dataset CDB90G are subdivided into various battles in the dataset. There are a total of 67 campaigns and 664 battles. The details of the battles are given in Table 10.

Table 10. Number of Battles per Campaigns

NUMBER OF BATTLES PER CAMPAIGNS	
Campaign	Number of Battles
American Civil War	49
American Revolution	14
Arab–Israeli War, 1967 (Six-Day War)	19
Arab–Israeli War, 1968	2
Arab–Israeli War, 1973 (October War)	33
Arab–Israeli, 1948	9
Arab–Israeli, 1956	4
Arab–Israeli, 1967	6
Austro–Italian War	1
Austro–Prussian (Seven Week's) War	1
Austro–Turkish War	4

Campaign	Number of Battles
Boer War	5
Crimean War	2
Dutch War	5
Egypt and Sudan	2
English Civil War	6
Franco–Prussian War	10
Franco–Spanish War	1
Great Northern War	1
Israel–Lebanon, 1982	1
Italo–Ethiopian War	1
Jacobite Rebellion	2
King William's War	8
Korean War	11
Latin American Wars of Independence	6
Monmouth's Rebellion	1
Netherland's War of Independence	1
Polish–Turkish War	1
Russo–Finish War (1939–1940)	1
Russo–Japanese War	6
Russo–Polish War (1920)	2
Second English Civil War	3
Spanish–American War	1
The Balkan Wars	5
The Fronde	1
The Manchurian Incident (1938–1939)	5
The Napoleonic Wars	29
The Seven Year's War	18
The Spanish Civil War	1
Thirty Year's War	18
Transvaal revolt	1
US–Mexican War	8
Vietnam	1
War of 1812	4
War of Austria with France and Piedmont	2
War of Texan Independence	1
War of the Austrian Succession	7
War of the First Coalition	14
War of the Second Coalition	7
War of the Spanish Succession	4

Campaign	Number of Battles
World War I	30
World War I (Eastern front 1914)	9
World War I (Italian front 1915)	4
World War I (Serbian front 1914)	2
World War I (Turkish fronts 1915)	4
World War I (Turkish fronts 1917)	5
World War I (Western front 1914)	16
World War I (Western front 1918)	54
World War I	32
World War II (European Theater)	28
World War II (Italy 1943–1944)	40
World War II (Italy 1944)	24
World War II (North Africa 1942–1943)	8
World War II (Pacific, 1943–1945)	4
World War II, Eastern Front	29
World War II, Okinawa	28
Zulu War	2

3. Battles According to Months

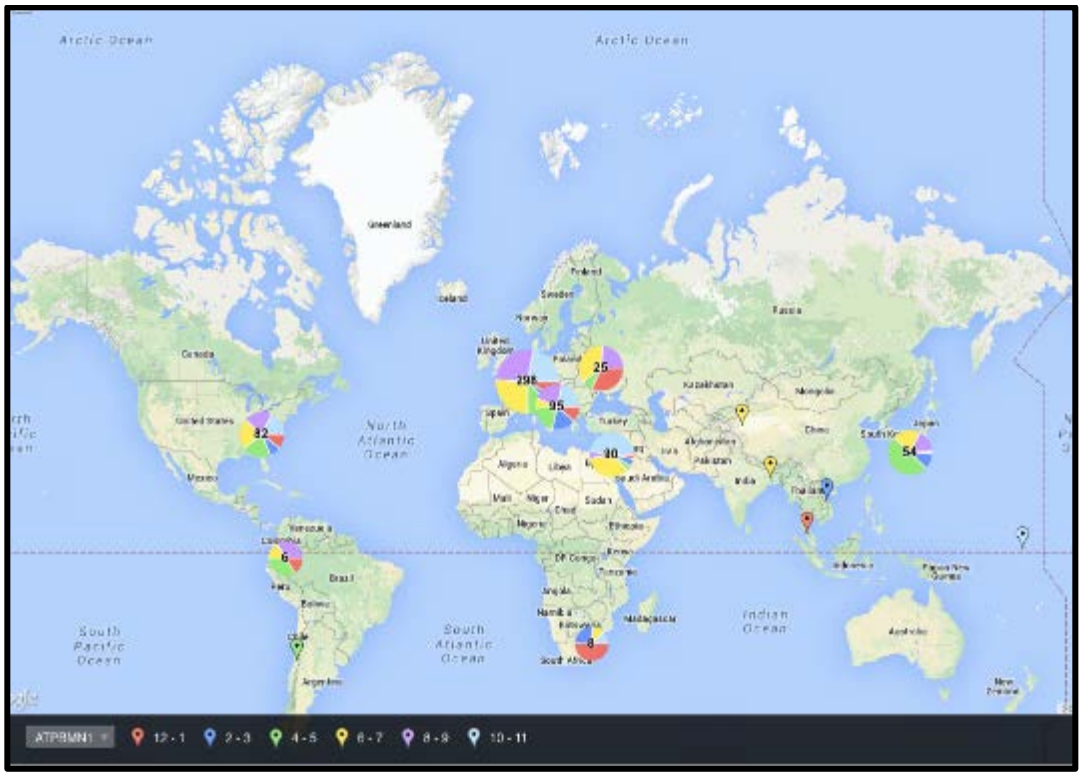
Sometimes climatic conditions played an important role in initiating battles. It is a well-documented fact that campaigns in Russia during winter favored the Russians. Likewise, areas that experienced relatively hot summers carried out military campaigns in winter. Based on this phenomenon, the dataset was explored to determine the frequency of different months in which battles were initiated (see Table 11). The data show that most battles began in June, and the fewest began in March.

Table 11. Number of Battles in Different Months

NUMBER OF BATTLES IN DIFFERENT MONTHS	
Months	Number of Battles
January	28
February	27
March	25
April	41
May	68
June	95
July	62
August	72
September	68
October	89
November	53
December	36

These data help in visualizing the battle areas in accordance with campaign seasons. Figure 7 depicts battles from December–January, February–March, April–May, June–July, August–September and October–November. This figure shows that in the Middle East, most battles were fought in October–November, which are relatively cooler months in the region.

Figure 7. Battle Locations in Accordance with Different Months



4. Battle Outcome: "WINA"

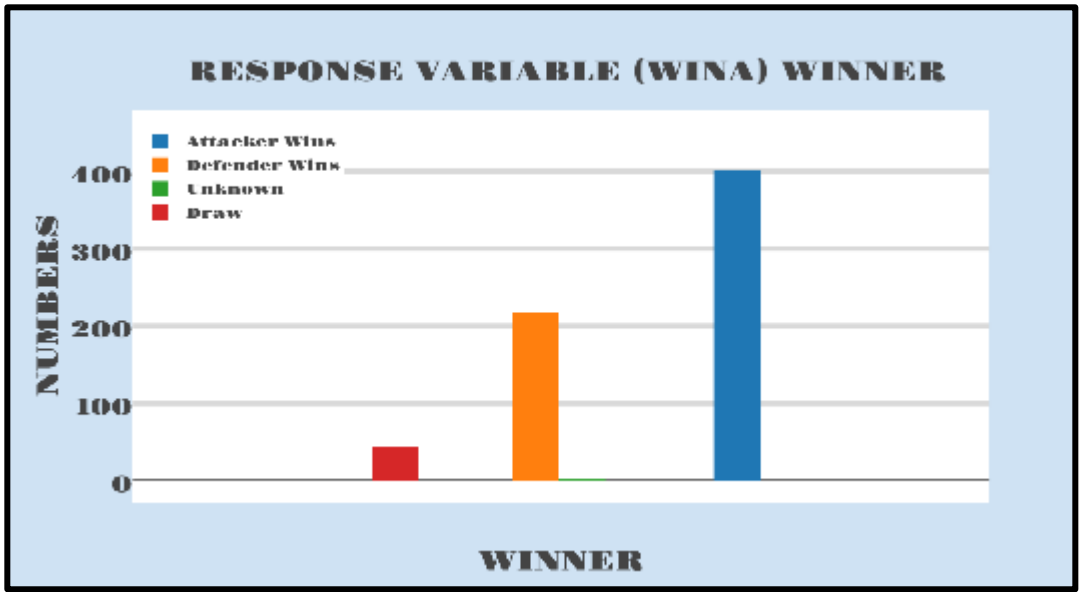
In every battle there comes a time when both sides consider themselves beaten, then he who continues the attack wins. – Ulysses S. Grant

The response variable WINA indicates when an attacker won, lost, or had a draw. According to the CDB90G dataset, the criteria was did the attacker achieve their objective. WINA has four levels: attacker wins (1), defender wins (−1), draw (0), and unknown (−9). Out of 664 battles, the attacker won 60.39%, the defender won 32.68%, and there was a draw 6.62% of the time. Table 13 and Figure 8 reflect the results of this response variable. The battles with unknown results are not modeled.

Table 12. Response Variable

RESPONSE VARIABLE			
Attacker Wins (1)	Defender Wins (-1)	Draw (0)	Unknown (-9)
401	217	44	2

Figure 8. Battles Won by Attacker/ Defender/ Draw or Unknown



5. Force Ratio

"God indeed sides with the biggest battalions." –*Ulysses S. Grant*

Force ratio is defined as the ratio of the attacker's personnel to the defender's personnel. It is assumed that force ratio is a valuable predictor of battle outcome. It is a common misconception that larger forces are typically more successful than smaller forces. However, in analyzing the history of war, it is apparent that force ratio is not necessarily a prerequisite for winning a war. Militaries around the world have adhered to an attacker to defender force ratio of 3:1. A statistical analysis of this force ratio is performed to examine whether the accepted 3:1 standard is justified. However, force ratio

alone cannot reliably predict the outcome of a battle. Force ratio's relationship with other tangible and intangible factors influences the outcome of battle, both directly and indirectly.

The exploration of force ratio provides insight into the 3:1 rule of thumb. Using the dataset, we created 13 categories of force ratios on a scale of 0.3 to 16.8. According to the data, there were 317 battles in which the ratio was less than 3:1 yet the attacker managed to win the battle. Based on this dataset, 69.4% of the time, the attacker wins when force ratio is greater than 3:1. The median force ratio for attacker wins is 2, and the median for defender win is 1. Table 15 gives a range of force ratios in the dataset along with the number of battles won by attackers and defenders.

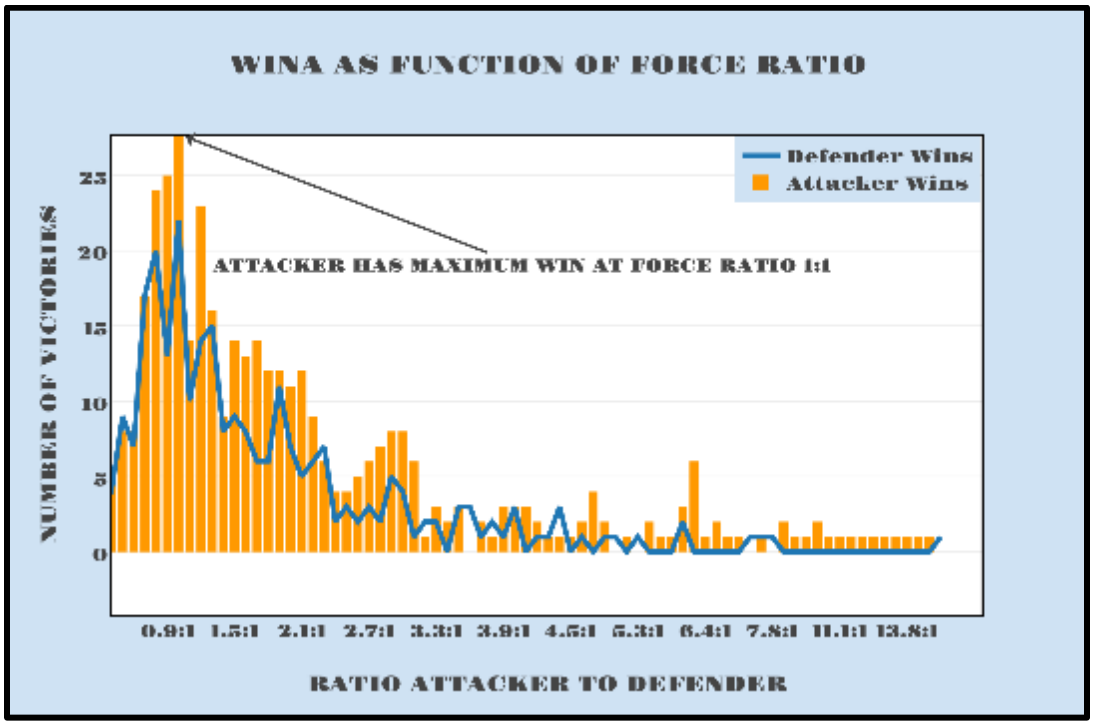
Table 13. Summary Statistics Force Ratio

FORCE RATIO		
Ratio	Attacker Wins (1)	Defender Wins (-1)
0.3:0.9	90	73
1.0:1.9	155	109
2.0: 2.9	72	42
3.0:3.9	29	19
4.0:4.9	19	11
5.0:5.9	8	3
6.3:6.9	10	0
7.0:7.9	4	3
8.2:8.5	2	0
9.5:9.6	3	0
11.1:11.9	4	0
12.0:12.4	2	0
13.8:16.8	3	1

Figure 9 reflects the comparison of force ratios and wins by the defender and attacker. The force ratios increase from left to right, and as force ratios increases the number of victories decreases. One reason for this is that a force ratio of 6:1 or higher is very rare. In this dataset of 662 battles, there were only 32 battles in which there was a

force ratio greater than 6:1. Of this group, the attackers won 28 battles, and the defenders won four.

Figure 9. Force Ratio Attacker to Defender



6. Tank Ratio

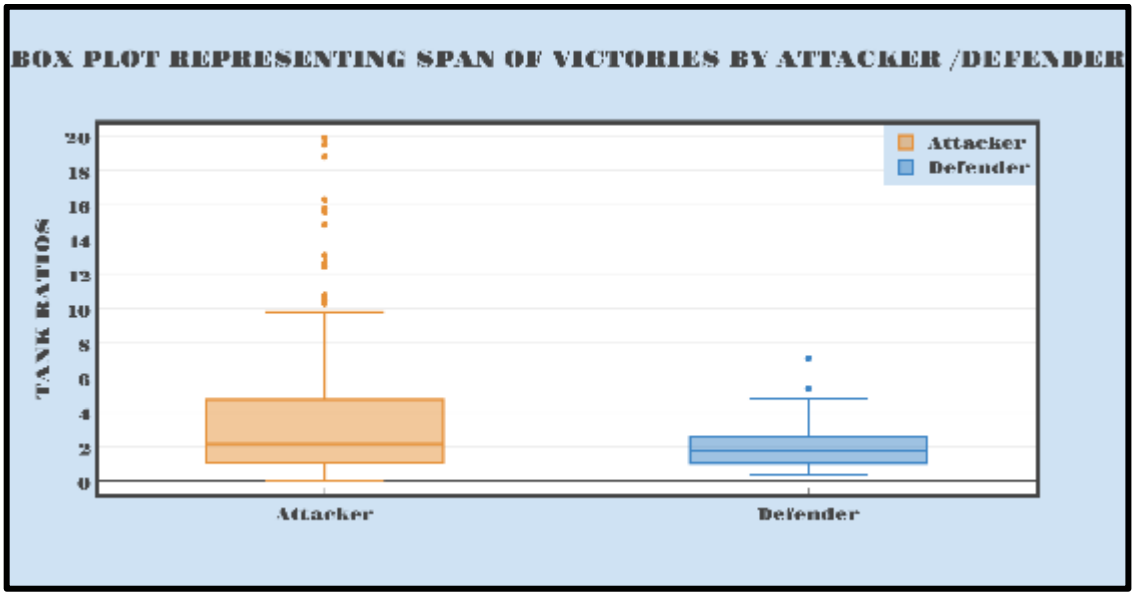
Tank ratio is defined as the ratio of the attacker's number of tanks to the defender's tanks. Tank Ratio is based on two variables: attacker tanks and defender tanks. Detailed statistics regarding the number of wins by attacker and defender depending upon tank use is given in Table 14. The dataset indicates 300 battles in which tanks were used. For the purpose of analysis, only battles in which both sides have tanks are considered.

Table 14. Summary Tank Metrics

SUMMAR TANK METRICS	
Total Number of Battles in which Tank were used by either side	300
Defender	
Number of Battles Defender Won Without Tanks when Attacker has Tanks	8
Number of Battles Defender Won and Attacker has no Tanks	5
Number of Battles Defender Won when both sides have Tanks	62
Total Number of Battles Won by Defender	75
Attacker	
Number of Battles Won By Attacker Without Tanks when Defender has tanks	8
Number of Battles Won By Attacker when Defender has no Tanks	56
Number of Battles Won by Attacker when both sides have Tanks	135
Total Number of Battles Won by Attacker	199
Draw	
Number of Battles Draw when both sides have Tanks	26

The tank ratio mean for when the attacker won is 3.75, the median is 2.16, and the standard deviation is 4.21. The tank ratio mean for when the defender won is 1.98, the median 1.79, and the standard deviation 1.33. Figure 10 is a box plot representing victories by each side. The box plot demonstrates that attackers have more victories than defenders with higher tank ratios.

Figure 10. Tank Ratio Attacker to Defender



7. Artillery Ratio

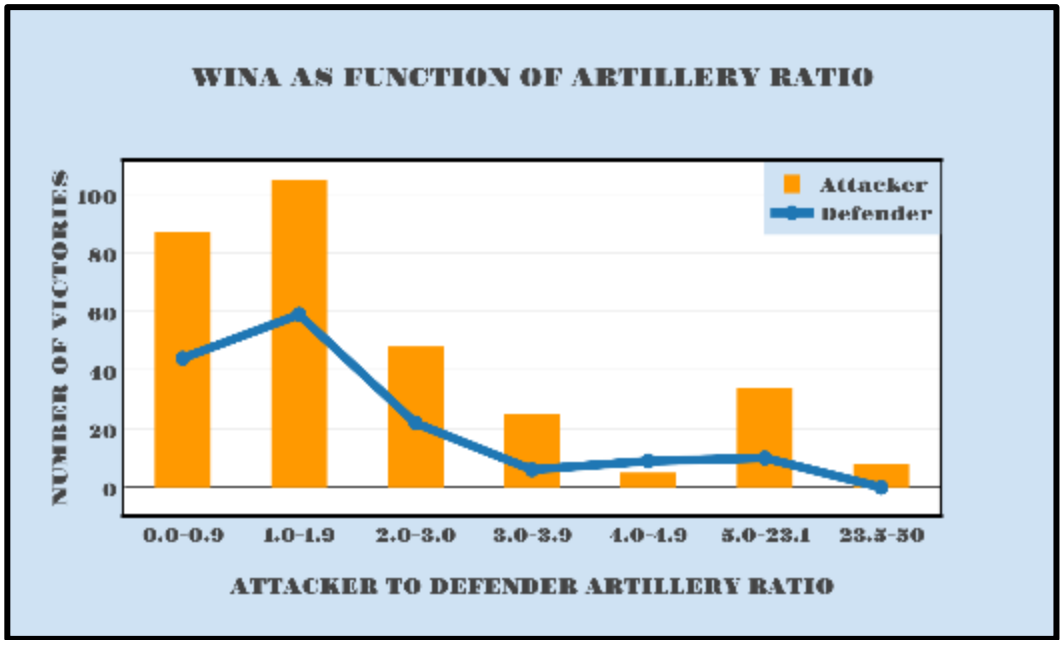
Artillery ratio is defined as the ratio of the attacker's number of artillery pieces to the defender's artillery pieces. Artillery has played an important role in warfare. Artillery is the only weapon that has been used from the first to the last battle in the dataset. Out of 664 battles, the strength of the attacker's artillery is unknown in 107 battles. The strength of the defender's artillery is unknown in 121 battles. In 32, neither side had artillery. Within the dataset, 492 battles exist in which both sides had artillery. Out of these, the defender won 146 battles, the attacker won 311, and 35 battles resulted in a draw.

Table 15. Summary Statistics Artillery

SUMMARY STATISTICS ARTILLERY ATTACKER TO DEFENDER	
Defender Wins Median	1.34
Defender Wins Mean	1.92
Defender Win Standard Deviation	2.17
Attacker Wins Median	1.51
Attacker Wins Mean	2.55
Attacker Win Standard Deviation	5.62

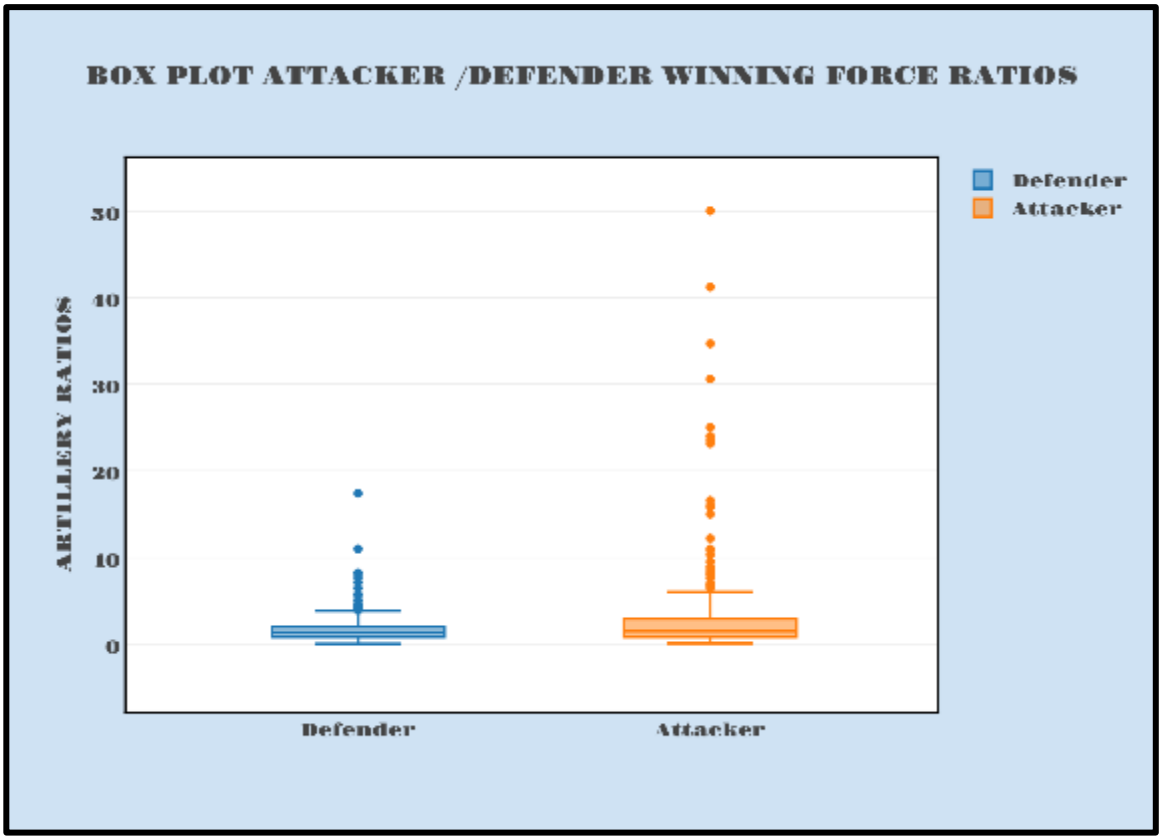
The graphical depiction of the artillery ratio is given in Figure 11. Most of the victories for both defenders and attackers occur when the artillery ratio is within a range of 1–1.09. As the artillery ratio increases in favor of the attacker, the number of victories for the defender decreases.

Figure 11. Artillery Ratio Attacker to Defender



The box plot of artillery ratio is given in Figure 12. The spreads of the attacker and the defender winning ratios are quite close; however, there are more outliers in attackers than defenders.

Figure 12. Artillery Ratio Attacker to Defender



8. Initial Force Ratio

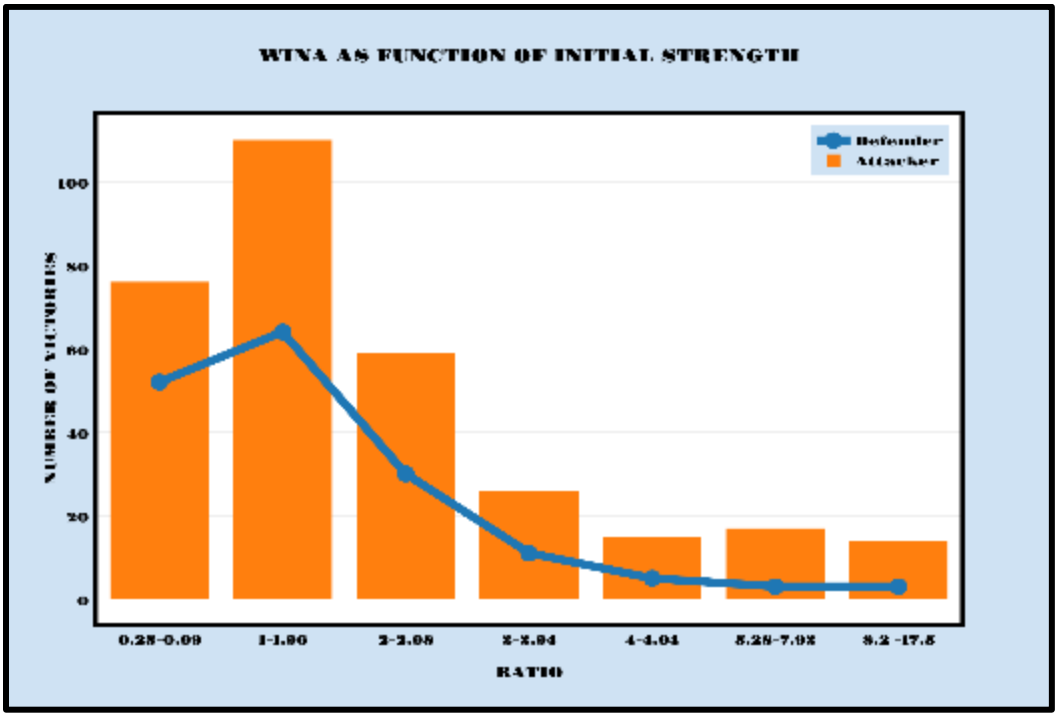
Force ratio is defined as the ratio of the attacker's initial number of personnel to the initial defender's number of personnel. Initial force ratio is another important variable that impacts the mass and offensive capabilities of the force. Theoretically, this ratio can be one of the major predictors for battle outcome. Initial force ratio is the force strength at the beginning of the battle. This variable requires a detailed analysis to determine its impact on winning battles. Detailed summary statistics, including the median, mean, and standard deviation are provided in Table 16.

Table 16. Summary Statistics Initial Force Ratio

SUMMARY STATISTICS INITIAL FORCE RATIO ATTACKER TO DEFENDER	
Defender Wins Median	1.31
Defender Wins Mean	1.92
Defender Win Standard Deviation	2.23
Attacker Wins Median	1.66
Attacker Wins Mean	2.43
Attacker Win Standard Deviation	5.47

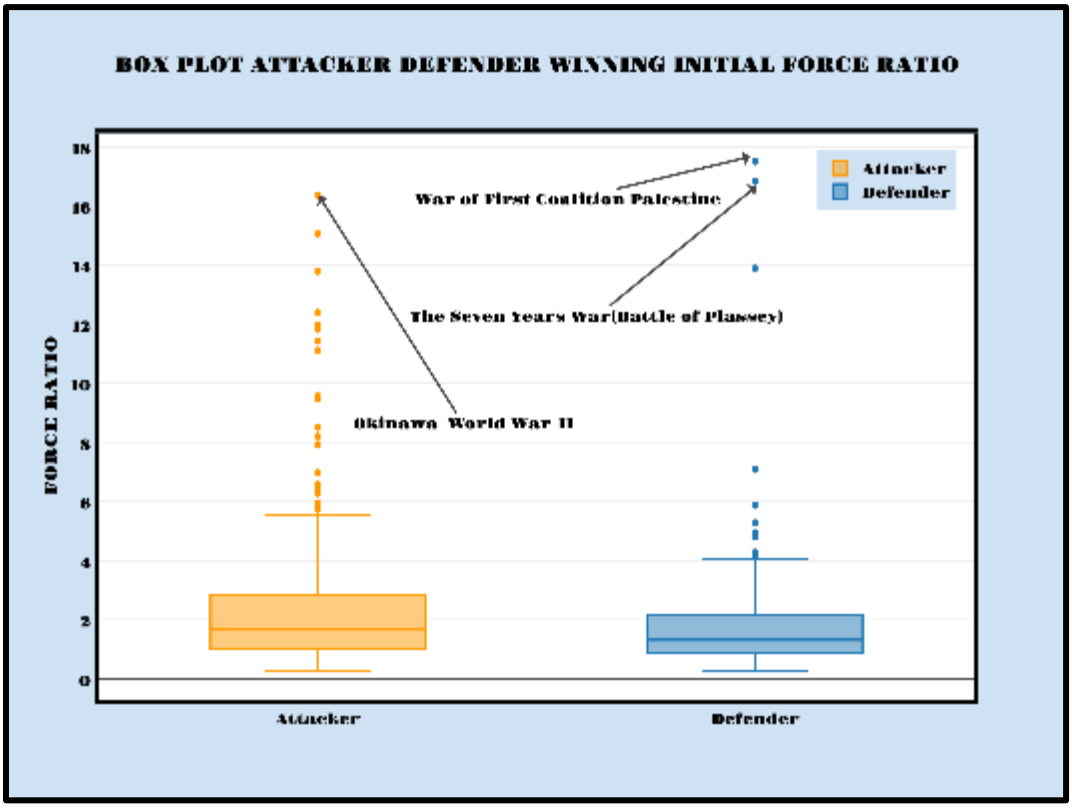
Figure 13 shows the number of victories by the defender and attacker as a function of the initial ratio. Most of the victories are within the range of 1–1.96. As the ratio in favor of the attacker increases, the number of victories by the defender decreases.

Figure 13. Initial Strength Attacker to Defender



The spread of defender and attacker winning ratios is given in Figure 16. The attacker spread is more than defender, and both have outliers as shown in Figure 14.

Figure 14. Initial Strength Attacker to Defender



9. Cavalry Ratio

Cavalry ratio is defined as the ratio of the attacker's number of horses to the defender's horses. Cavalry played an important role during the battles of the seventeenth and eighteenth centuries. However, with the development of motorized vehicles, cavalry started diminishing on battlefields. The dataset contains battles in which cavalry has been used (see Table 17).

Table 17. Cavalry Metrics

CAVALRY METRICS				
	Total	Attacker Wins	Defender Wins	Draw
Number of Battles With Cavalry	137	84	49	4
Number of Battles Without Cavalry on both side	423	266	121	36
Number of Battles with Unknown Cavalry	93	48	42	3
Number of Battles with No Cavalry on Attacker Side	2	1	1	0
Number of Battles with No Cavalry on Defender Side	7	2	4	1

Summary statistics for the cavalry ratio are given in Table 18. The winning ratio mean for the defender is greater than the attacker mean, whereas the winning median for the attacker is greater that of the defender.

Table 18. Summary Statistics Cavalry Ratio

SUMMARY STATISTICS CAVALRY RATIO ATTACKER TO DEFENDER	
Defender Wins Median	1.11
Defender Wins Mean	1.51
Defender Win Standard Deviation	1.27
Attacker Wins Median	1
Attacker Wins Mean	1.29
Attacker Win Standard Deviation	1.25

As shown in Figure 15, most victories for the attacker fall within the range of 0.1–0.97, and most victories for the defender fall within the range of 1–1.95.

Figure 15. WINA as a Function of Cavalry Ratio

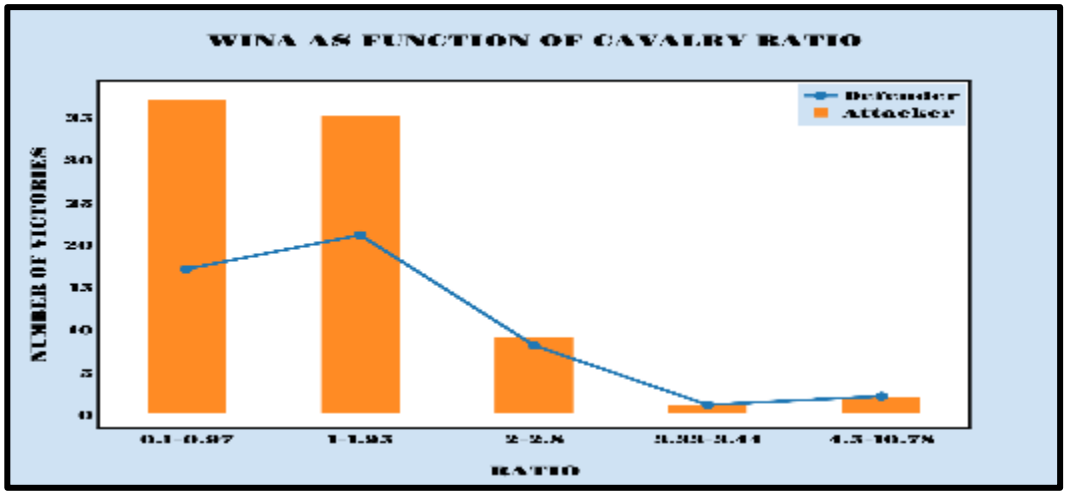
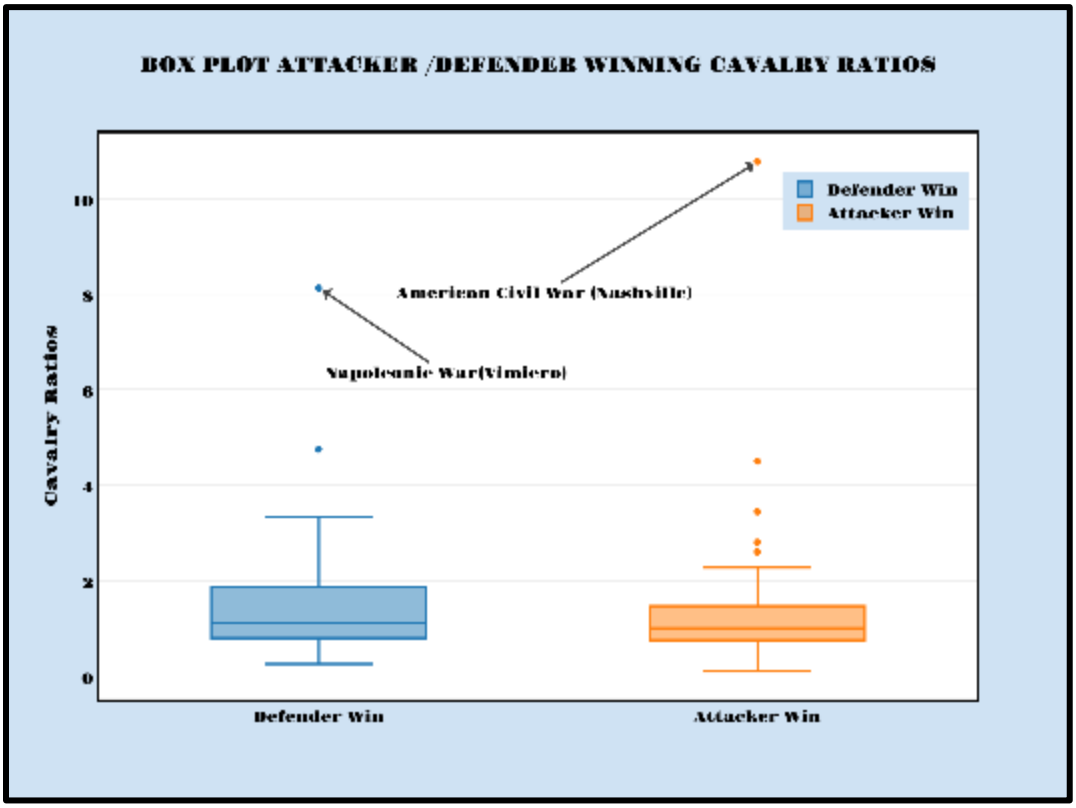


Figure 16 gives a spread of the winning cavalry ratios for both attackers and defenders. Defenders have more spread than attackers. However, attackers have more outliers than do the defenders.

Figure 16. Box Plot Winning Cavalry Ratio



10. Close Air Support

The earliest known form of air support is from the Chinese, who used manned kites for spying and communication. Around the second or third century, the Kongming lantern was invented by the Chinese to use for military communication. In 1794, the French Aerostatic Corps at the battle of Fleurus used a tethered balloon to gain a vantage point. In 1912, against the Turks at Ain Zara, Italians first used airplanes in battle. Greek aviators performed the first naval joint air operation in history during the Dardanelles campaign (1915–1916). In the dataset, there were 88 battles in which both sides had air support available. There were 332 battles in which both sides had no air support available (see Table 19).

Table 19. Close Air Support Ratio Metrics

CLOSE AIR SUPPORT RATIO METRICS				
	Total	Attacker Wins	Defender Wins	Draw
Number of Battles With Close Air Support	88	49	35	4
Number of Battles Without Close Air Support on both side	332	188	125	19
Number of Battles with Unknown Close Air Support	132	87	35	10
Number of Battles with No Close Air Support on Attacker Side	342	191	131	20
Number of Battles with No Close Air Support on Defender Side	418	250	140	28

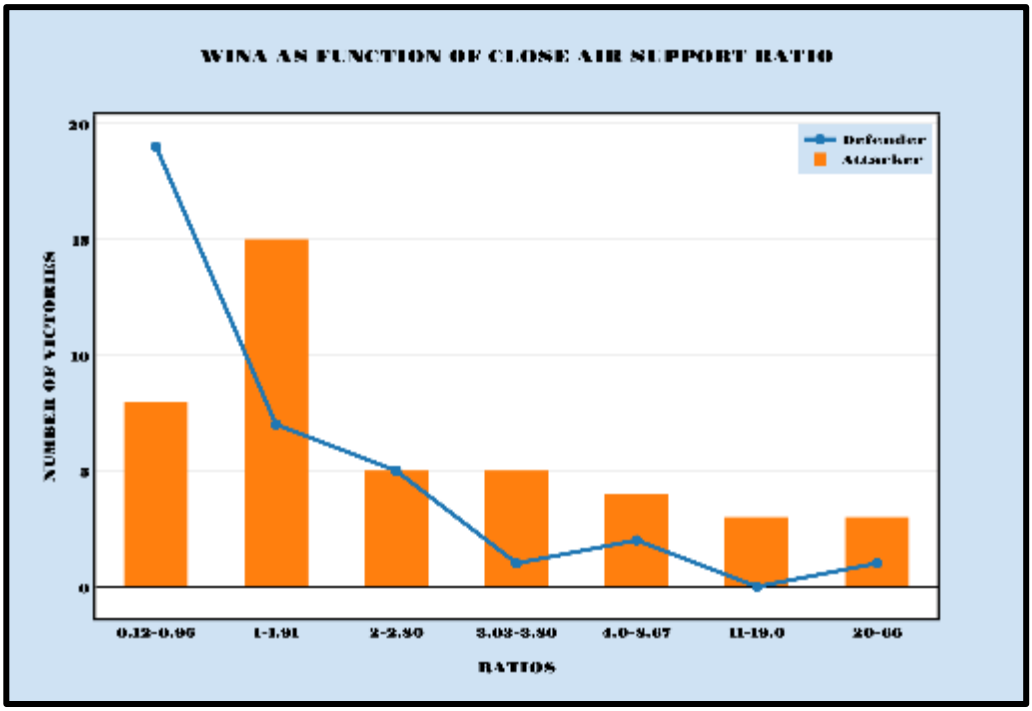
Summary statistics for the air support ratio is given in Table 20. The defender's air support mean, median, and standard deviation are less than those of the attacker.

Table 20. Summary Statistics Air Support Ratio

SUMMARY STATISTICS AIR SUPPORT RATIO	
Defender Wins Median	0.94
Defender Wins Mean	2.24
Defender Win Standard Deviation	5.65
Attacker Wins Median	2.05
Attacker Wins Mean	5.11
Attacker Win Standard Deviation	10.45

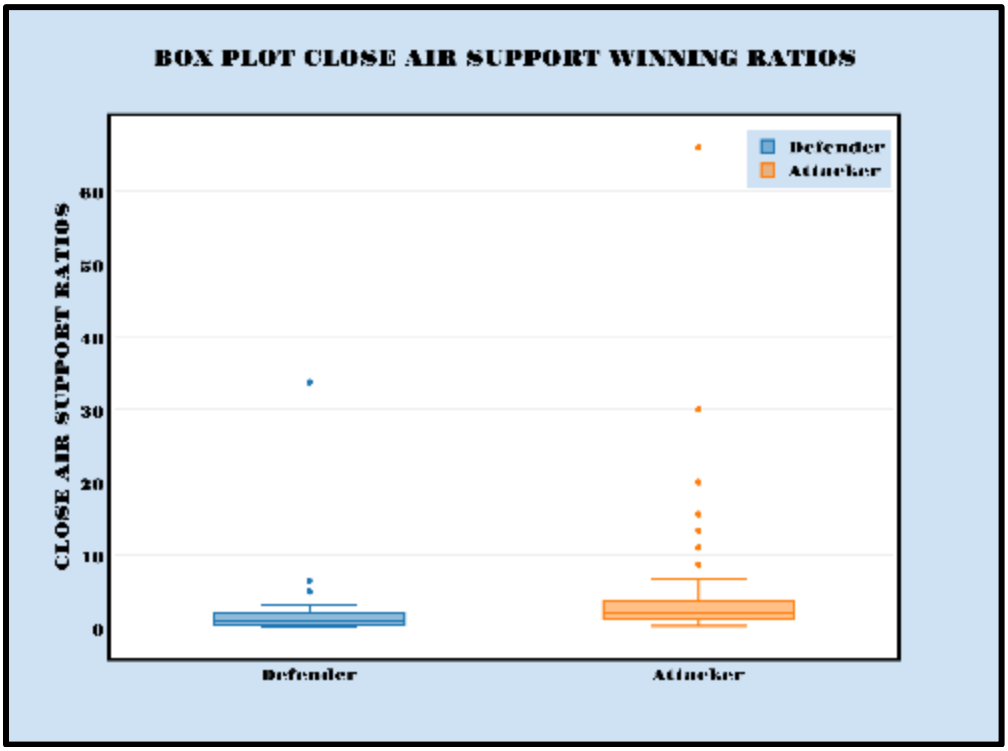
Figure 17 indicates that most of the victories for the attacker fall within the range of 1–1.91, and most victories for the defender fall within ratio of 0.12–0.96. Within the range of 0.12–0.96, the defender has more victory. This means that close air support matters and plays a role in victory. As air support ratio of the attacker increases, the number of victories for the defender decreases.

Figure 17. WINA as a Function of Close Air Support Ratio



The box plot in Figure 18 shows that the attacker mean is greater than the defender mean. The spread of the attacker's winning ratio is much more than that of the defender. There are more outliers for attackers then there are for defenders. One of the most significant outliers was the battle of Lariano, where the attacker to defender air support ratio was 66:1.

Figure 18. Box Plot Close Air Support Winning Ratios



11. Attackers Primary Tactical Scheme

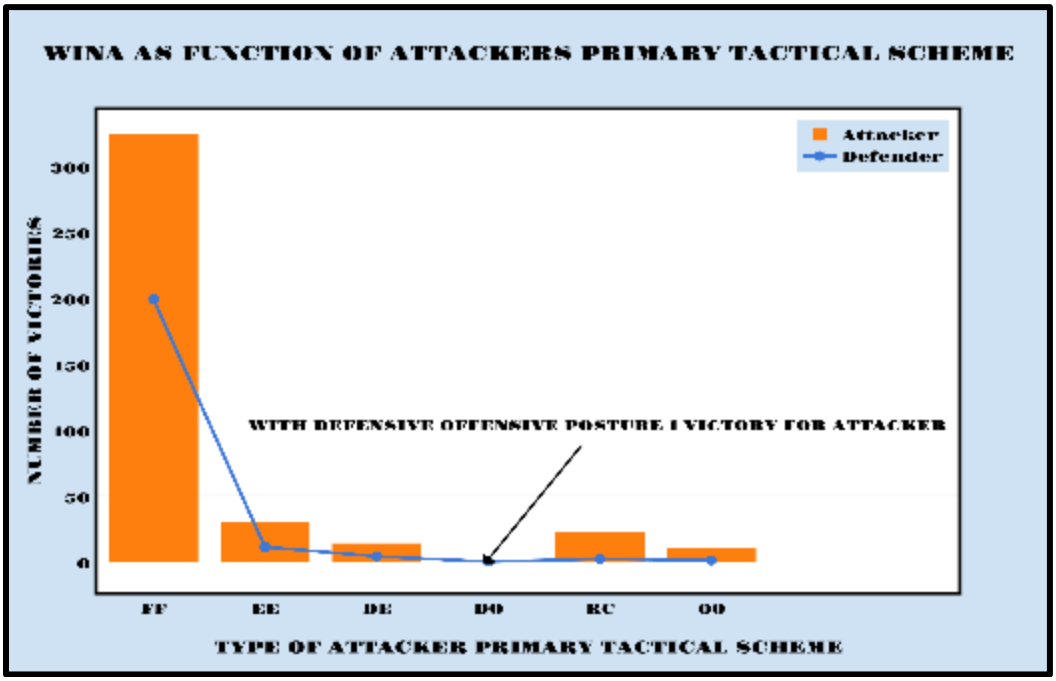
The attacker primary tactical scheme variable connects with the principles of offense and movement. In military terminology, maneuvering is the ways in which attackers strike the defender to achieve their objective. Throughout military history, armies developed attack techniques to achieve their objectives. These techniques have varied based on mission, force, terrain, and availability of resources. The variable of attackers primary tactical scheme given in the codebook for this dataset has 13 different types. However, information on six types is available in the dataset. Details about each of these variables is outlined in Table 21.

Table 21. Attacker Primary Tactical Scheme Metrics

ATTACKERS PRIMARY TACTICAL SCHEME				
	Total	Attacker Wins	Defender Wins	Draw
Frontal Attack (FF)	563	324	199	40
Single Envelopment (EE)	41	30	11	0
Double Envelopment (DE)	19	14	4	1
Defensive /offensive plan (DO)	1	1	0	0
River Crossing (RC)	27	22	2	3
Not Applicable / Not Available (OO)	11	10	1	0

The frontal attack is the most common technique used by the attacker. Of the database of 662 battles, 563 battles involved frontal attacks. One of the reasons for this is the simplicity of the technique as compared to other methods. Figure 19 compares the different techniques and the number of victories by technique.

Figure 19. Attackers Primary Tactical Scheme



12. Defender Primary Posture

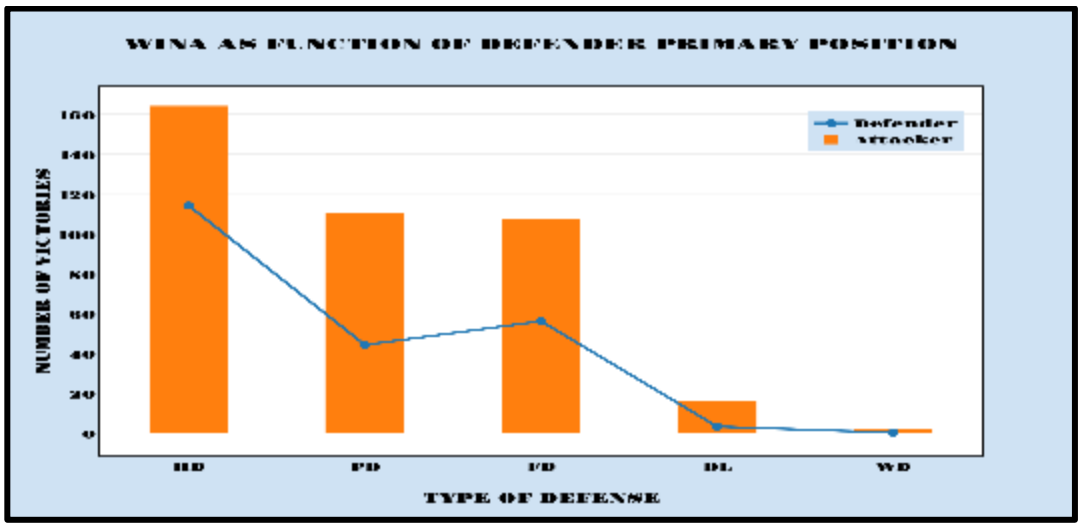
The variable *defender primary posture* defines the type of defensive position adopted by the defender. The variable has six levels; an overview of defender primary posture is given in Table 22.

Table 22. Defender Primary Posture Metrics

DEFENDER PRIMARY POSTURE (POST1)				
	Total	Attacker Wins	Defender Wins	Draw
Hasty Defense (HD)	293	164	114	15
Prepared Defense (PD)	166	110	44	12
Fortified Defense (FD)	179	107	56	16
Delaying Action Adopted (DL)	20	16	3	1
Withdrawal Adopted (WD)	2	2	0	0
Not Applicable / Not Available (OO)	2	2	0	0

Figure 20 shows different types of defensive positions and victories by both attackers and defenders. Most of these attacks were carried out against a hasty defense.

Figure 20. Defender Primary Posture



13. Attacker Primary Tactics Versus Defender Primary Posture

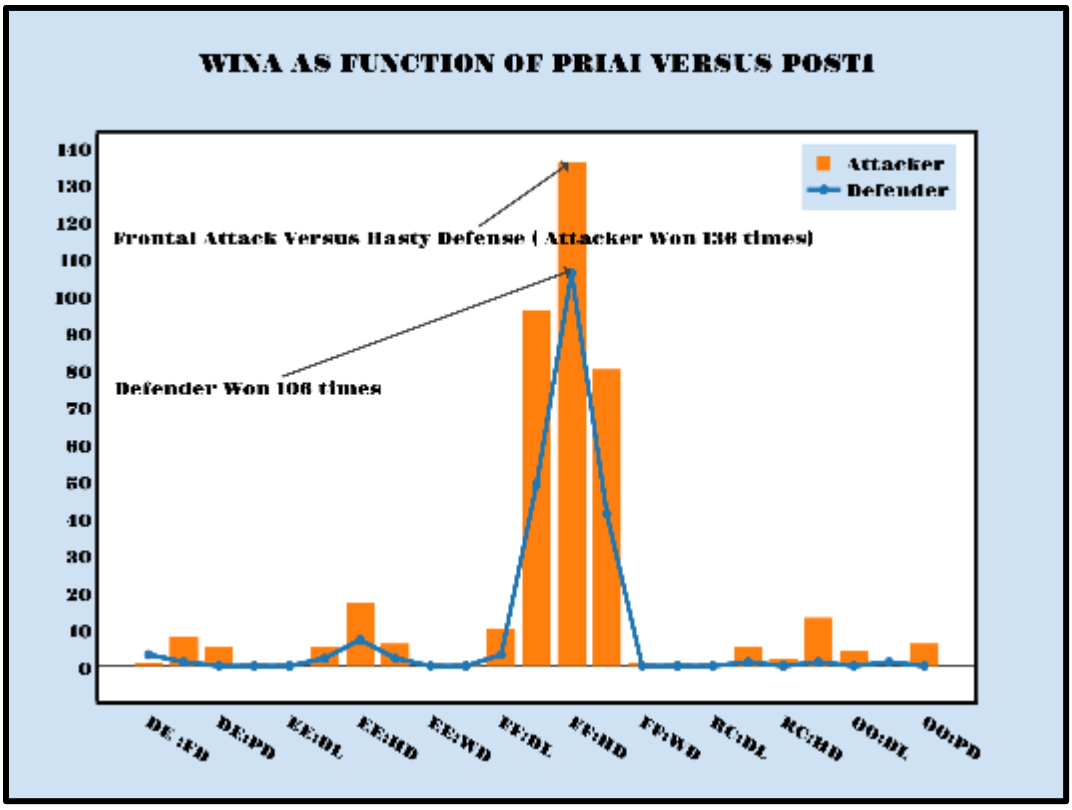
The *attacker primary tactics* and *defender posture* variables cannot be explained separately because they are proven to be correlated. In Table 23, a detailed summary of attacker tactics against defenders lists victories for both sides. The most common tactic used by the attacker was a frontal attack against a hasty defense. The second most common technique was a frontal attack against a fortified defense. Fifteen observations in the dataset contain river crossing against a prepared defensive position. Out of those 15 times, the attacker achieved success 13 times in overcoming the prepared defensive position after carrying out a river-crossing.

Table 23. PRIAI Versus POST1

PRIAI VERSUS POST 1				
Maneuver Scheme		WINA (Number of Success)		
PRIAI	POST1	Attacker	Defender	Draw
DE	FD	1	3	0
DE	HD	8	1	1
DE	PD	5	0	0
DO	HD	1	0	0
EE	DL	1	0	0
EE	FD	5	2	0
EE	HD	17	7	0
EE	PD	6	2	0
EE	WD	1	0	0
FF	OO	1	0	0
FF	DL	10	3	0
FF	FD	96	49	15
FF	HD	136	106	14
FF	PD	80	41	11
FF	WD	1	0	0
RC	OO	1	0	0
RC	DL	1	0	1
RC	FD	5	1	1
RC	HD	2	0	0
RC	PD	13	1	1
OO	DL	4	0	0
OO	FD	0	1	0
OO	PD	6	0	0

Figure 21 gives an overview of the complete dataset of tactics employed by the attacker against the defensive postures. This is militarily intuitive; if the defender takes on a hasty defense, an attacker will not waste time in carrying out a difficult maneuver. Rather, they benefit by attacking the ill-prepared defensive force. Another noticeable trend is that a frontal attack is normally carried out against a hasty defense, withdrawal, and delaying action adopted by the defender.

Figure 21. WINA as a Function of Attacker Primary Tactics Versus Defender Primary Posture



14. Momentum Advantage

Measuring qualitative factors is difficult, and a scale of measurement differs according to criteria set by the organization compiling data. Momentum is also one of the qualitative criteria in the dataset. The compiling organization does not provide detailed criteria of measurement for momentum.

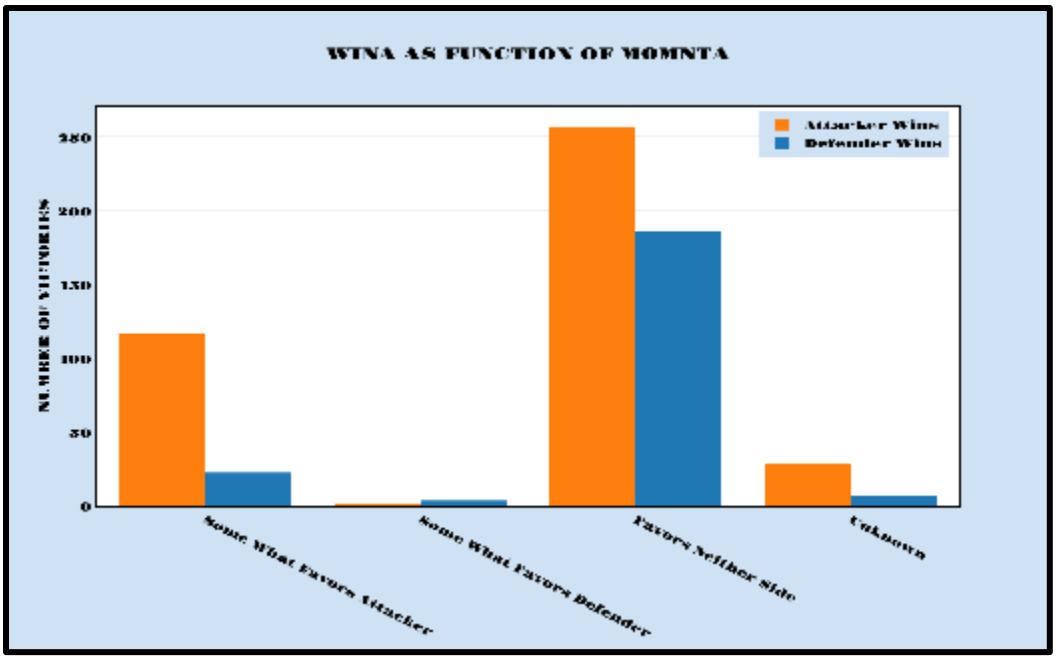
The relative momentum advantage variable can be correlated with the principles of offense, mass, and maneuver. Momentum advantage has 10 different levels, from very strongly favoring the attacker to unknown. Table 24 gives a detailed overview of momentum advantage in the CDB90G dataset.

Table 24. Relative Momentum Advantage Metrics

RELATIVE MOMENTUM ADVANTAGE (MOMNTA)				
	Total	Attacker Wins	Defender Wins	Draw
Somewhat favors the attacker	147	116	22	9
Somewhat favors the defender	4	1	3	0
Favors neither side	475	256	186	33
Unknown	36	28	6	2

Figure 22 indicates WINA as the function of momentum advantage. According to the dataset, most of the battles were fought with neither side having a momentum advantage over the other. However, in battles where the attacker held the advantage, the attacker achieved a high proportion of victories. Another noticeable fact is that the defender rarely holds momentum advantage. From a military perspective, this is due to the static nature of defensive warfare.

Figure 22. WINA as a Function of Relative Momentum Advantage



15. Air Superiority in Theatre

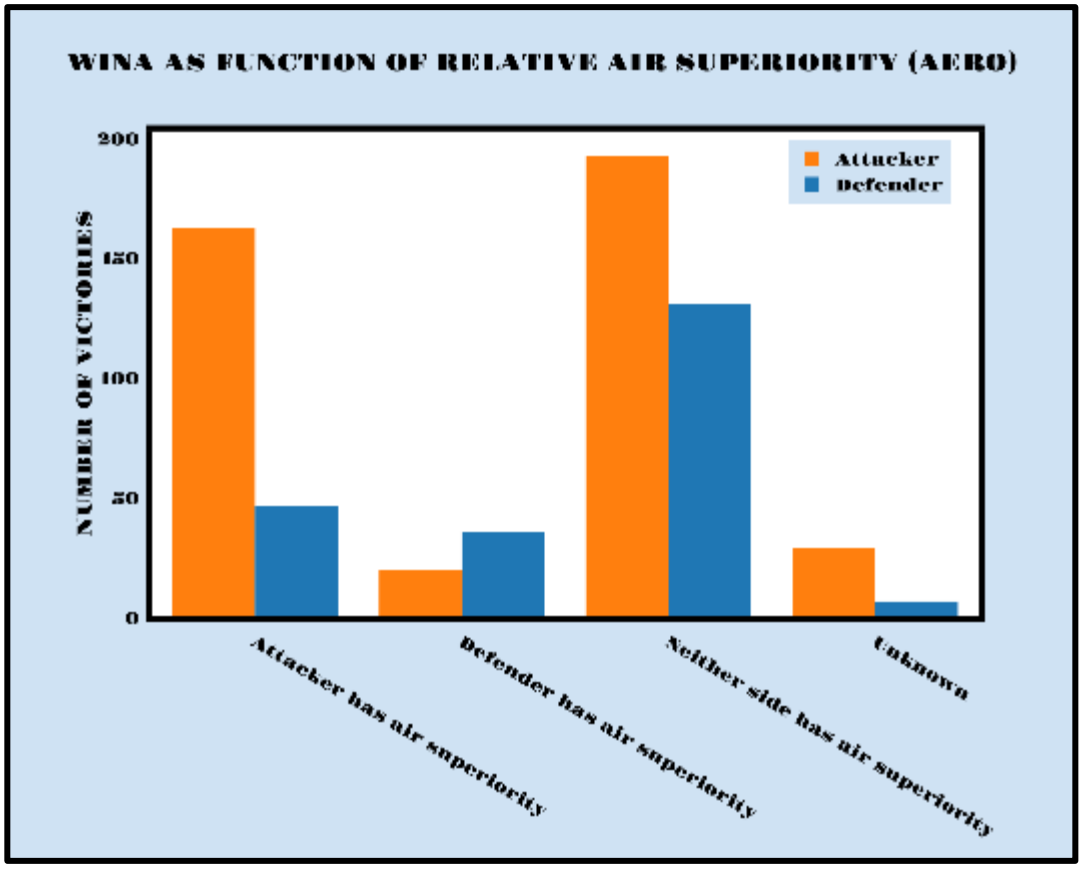
The relative air superiority variable contributes to mass and offensive principles. Aerial bombardment contributes significantly to the massing of firepower at particular points. Air superiority plays a vital role in the outcome of battle. However, in the dataset, there were 339 battles in which neither side had air superiority. Achieving air superiority requires tremendous resources. As shown in Table 25, there are 230 battles in which the attacker has air superiority and 57 battles in which the defender has air superiority.

Table 25. Relative Air Superiority in Theatre Metrics

RELATIVE AIR SUPERIORITY IN THEATRE				
	Total	Attacker Wins	Defender Wins	Draw
Attacker has air superiority	230	162	46	22
Defender has air superiority	57	19	35	3
Neither side has air superiority	339	192	130	17
Unknown	36	28	6	2

As indicated in Figure 23, most battles were fought without air superiority on either side. However, whenever the attacker had air superiority, the attacker usually won the battle. When the defender had air superiority, they usually won.

Figure 23. WINA as a Function of Relative Air Superiority



16. Initiative Advantage

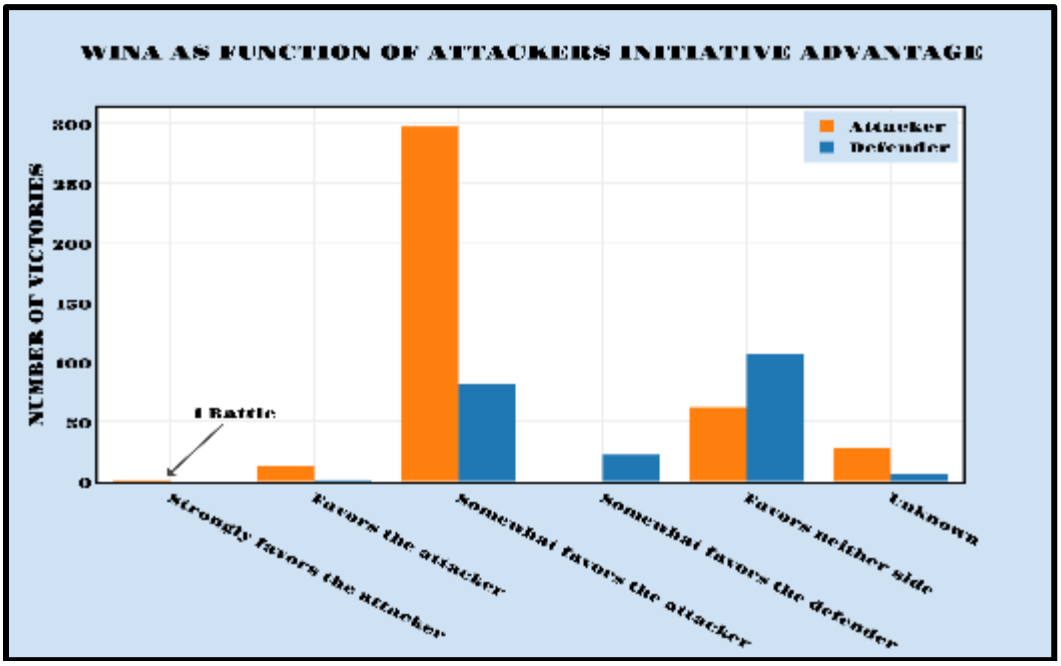
Initiative, like momentum, is another important variable that is difficult to measure and scale depending upon the criterion of measurement. The attacker usually holds a greater relative advantage than the defenders due to the attacker's ability to initiate hostilities. The dataset has defined this initiative with 10 different levels, details of which are given in the codebook for the dataset. Table 26 provides an overview of the attacker's relative advantage in relationship to victories gained.

Table 26. Relative Initiative Advantage Metrics

INITATIVE ADVANTAGE				
	Total	Attacker Wins	Defender Wins	Draw
Strongly favors the attacker	1	1	0	0
Favors the attacker	14	13	1	0
Somewhat favors the attacker	402	297	81	24
Somewhat favors the defender	24	0	22	2
Favors neither side	185	62	107	16
Unknown	36	28	6	2

Figure 24 indicates that if the attacker holds a relative initiative advantage, the attacker will usually succeed. However, in the complete dataset there is only one event in which relative initiative strongly favors the attacker. If the defender enjoys the relative initiative, the chances of success for the attacker are smaller.

Figure 24. WINA as a Function of Attackers Initiative Advantage



17. Surprise

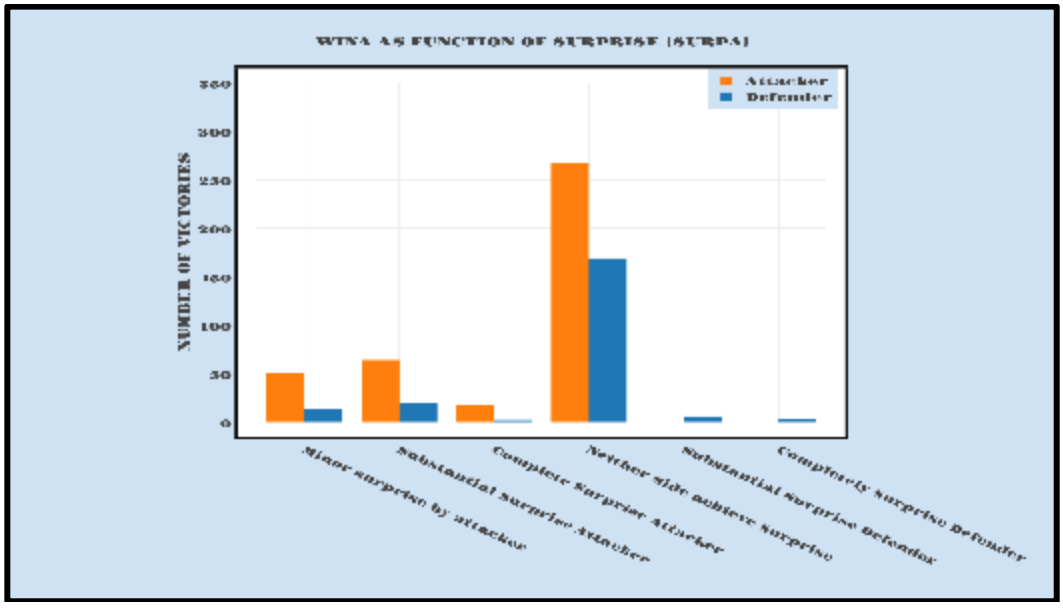
The surprise variable correlates with two principles of war: surprise and security. Because a force cannot achieve surprise without ensuring security of information, surprise is considered one of the most vital elements in winning a battle.

Table 27. Surprise Metrics

SURPRISE METRICS (SURPA)				
	Total	Attacker Wins	Defender Wins	Draw
Minor surprise by attacker	64	50	13	1
Substantial surprise by attacker	87	64	20	3
Complete surprise by attacker	20	18	2	0
Minor surprise achieve by defender	6	1	4	1
Substantial surprise by defender	6	0	6	0
Completely surprise by defender	3	0	3	0
Neither side achieves surprise	476	268	169	39

Data indicate that in most battles, both sides were not able to achieve surprise (see Figure 25). However, in case either side achieves surprise, it is usually succesful.

Figure 25. WINA as a Function of Surprise



18. Intelligence Advantage

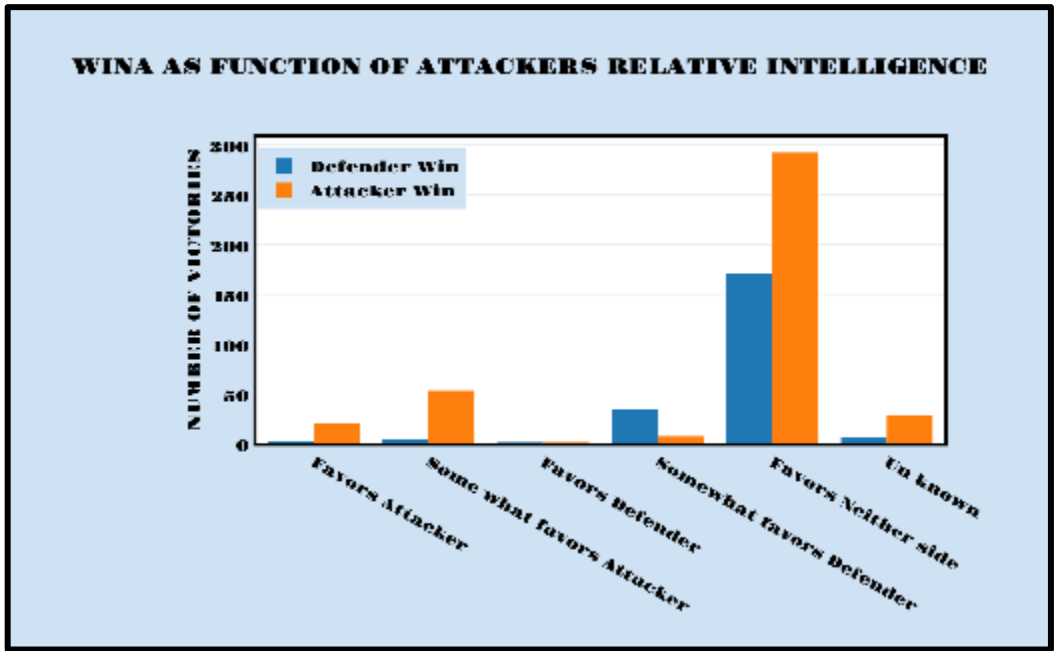
Intelligence, in military terminology, is considered as the capability to acquire information regarding the enemy. Intelligence has an important role in planning for battle since accurate information regarding the enemy can assist in developing good war plans. Intelligence is essential for security and surprise. However, surprise requires intelligence. Moreover, accurate intelligence usually ensures successful completion of the mission (see Table 28).

Table 28. Intelligence Advantage

INTELLIGENCE ADVANTAGE				
	Total	Attacker Wins	Defender Wins	Draw
Favors the attacker	22	20	2	0
Somewhat favors the attacker	58	53	4	1
Favors the defender	2	1	1	0
Somewhat favors the defender	46	7	34	5
Favors neither side	498	292	170	36
Unknown	36	28	6	2

Figure 26 indicates that victory is associated with an attacker's relative intelligence advantage. Usually, when the attacker has a relative intelligence advantage, it succeeds. When there is relative intelligence which favors the attacker, then the attacker wins 53 out of 58 battles. A similar trend can also be observed for the defender.

Figure 26. WINA as a Function of Relative Intelligence Advantage



19. Leadership Advantage

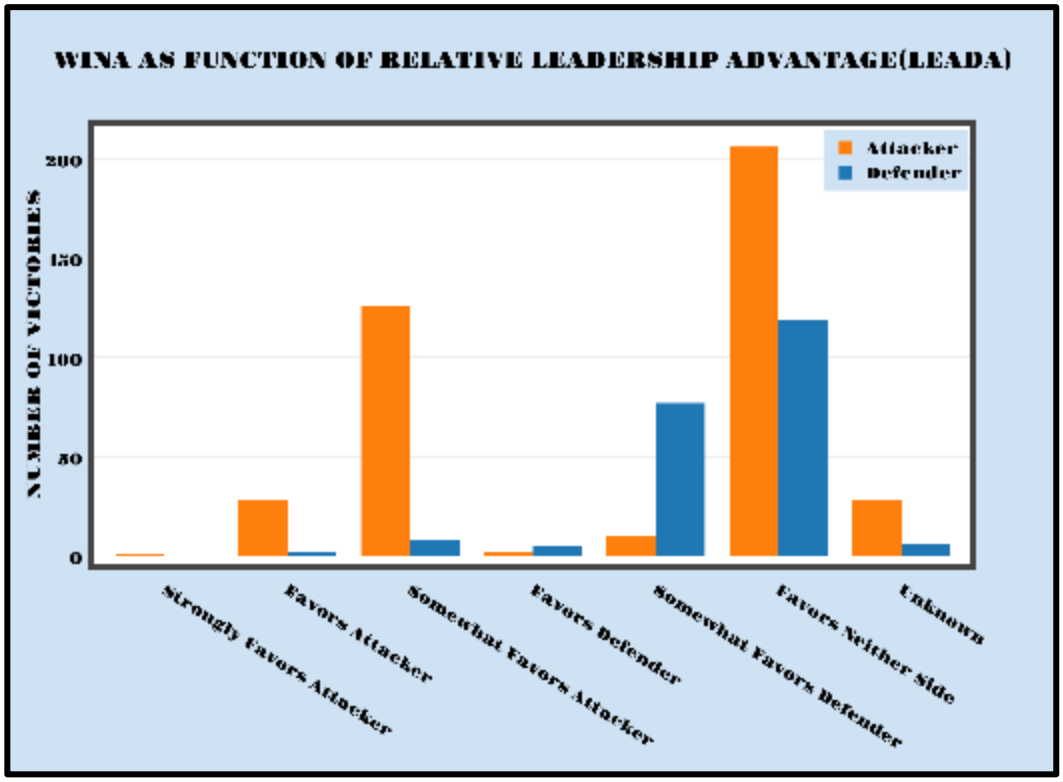
Measuring leadership is highly subjective and likely known only after the battles occur. Leadership has remained as an important variable in history. Charismatic leaders such as Alexander the Great, Hannibal, Erwin Rommel, and George Patton have played decisive roles in securing victories for their countries. In the dataset, the scale of measuring leadership advantage is from *strongly favoring the attacker* to *favors neither side*. A detailed description of these factors is given in Table 29. The table shows that whenever attackers or defenders hold a leadership advantage, they secure more victories. There are 30 battles in which the relative leadership advantage favors the attacker. Out of these, 28 attackers won. Similarly, there are 93 battles in which relative leadership advantage somewhat favors the defender. Out of 93, the defender won 77.

Table 29. Relative Leadership Advantage

RELATIVE LEADERSHIP ADVANTAGE				
	Total	Attacker Wins	Defender Wins	Draw
Strongly favors the attacker	1	1	0	0
Favors the attacker	30	28	2	0
Somewhat favors the attacker	137	126	8	3
Favors the defender	7	2	5	0
Somewhat favors the defender	93	10	77	6
Favors neither side	358	206	119	33
Unknown	36	28	6	2

Figure 29 indicates a relative leadership advantage. In most battles, neither side had a relative leadership advantage. When leadership advantage somewhat favors the attacker, most of the battles are won by the attacker.

Figure 27. WINA as a Function of Relative Leadership Advantage



20. Combat Effectiveness Advantage

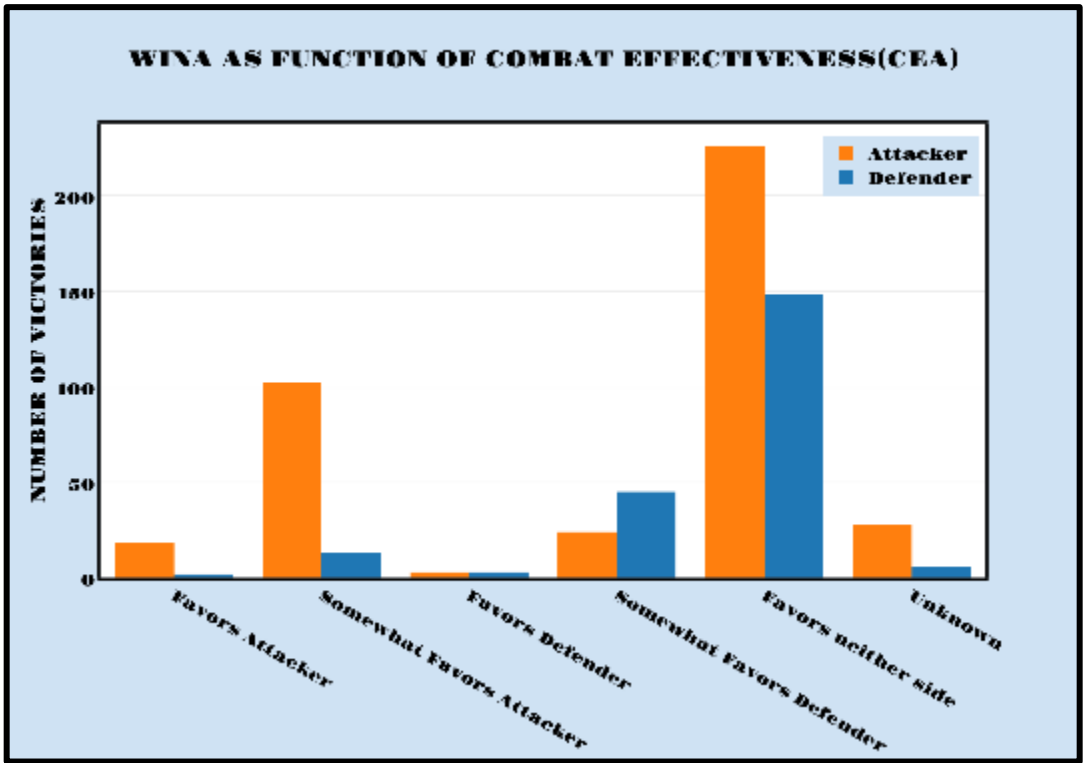
The combat effectiveness advantage is defined as combat efficiency advantage one side has over another. Like other relative variables combat efficiency,is difficult to measure and mostly dependent upon how historians define it. The combat effectiveness advantage seems to be an important variable based on the results displayed in Table 30. There are 20 battles in which combat effectiveness favors the attacker; of these 20 battles, the attacker won 18. The defender won 45 out of 76 battles in which combat effectiveness somewhat favors the defender.

Table 30. Combat Effectiveness

COMBAT EFFECTIVENESS ADVANTAGE				
	Total	Attacker Wins	Defender Wins	Draw
Favors the attacker	20	18	2	0
Somewhat Favors the attacker	123	102	13	8
Favors the defender	6	3	3	0
Somewhat favors the defender	76	24	45	7
Favors neither side	401	226	148	27
Unknown	36	28	6	2

Figure 28 reflects that most of the battles are fought when neither side has an advantage. In cases where relative combat effectiveness favors a side, then that side has more victories. According to the information in Figure 28, combat effectiveness tends to play an important role in winning the battle. However, in most of the battles, achieving a combat effectiveness advantage is difficult.

Figure 28. WINA as a Function of Combat Effectiveness



21. Logistic Advantage

Logistics in the military means a method for efficiently moving resources for combat. Logistics has played an important role in determining the outcome of many battles. In the dataset, most of the battles are fought without having any logistical advantage on one side (see Table 31). The dataset also indicates that in a case where each side has a logistical advantage over another, the chance of success for the side having the logistical advantage increases.

Table 31. Relative Logistic Advantage

RELATIVE LOGISTIC ADVANTAGE				
	Total	Attacker Wins	Defender Wins	Draw
Favors the attacker	11	10	1	0
Somewhat Favors the attacker	42	37	3	2
Favors the defender	1	0	0	1
Somewhat favors the defender	22	7	13	2
Favors neither side	550	319	193	38
Unknown	36	28	6	2

22. Training Advantage

All great leaders have trained their armies to achieve success in battles. The best-trained armies were most proficient in handling the rigors of war (see Table 32). This variable has a greater impact on the attacker than the defender. There are 13 battles in which training advantage favors the attacker. The result of these 13 battles overwhelmingly supports the attackers who win 12 out 13 battles.

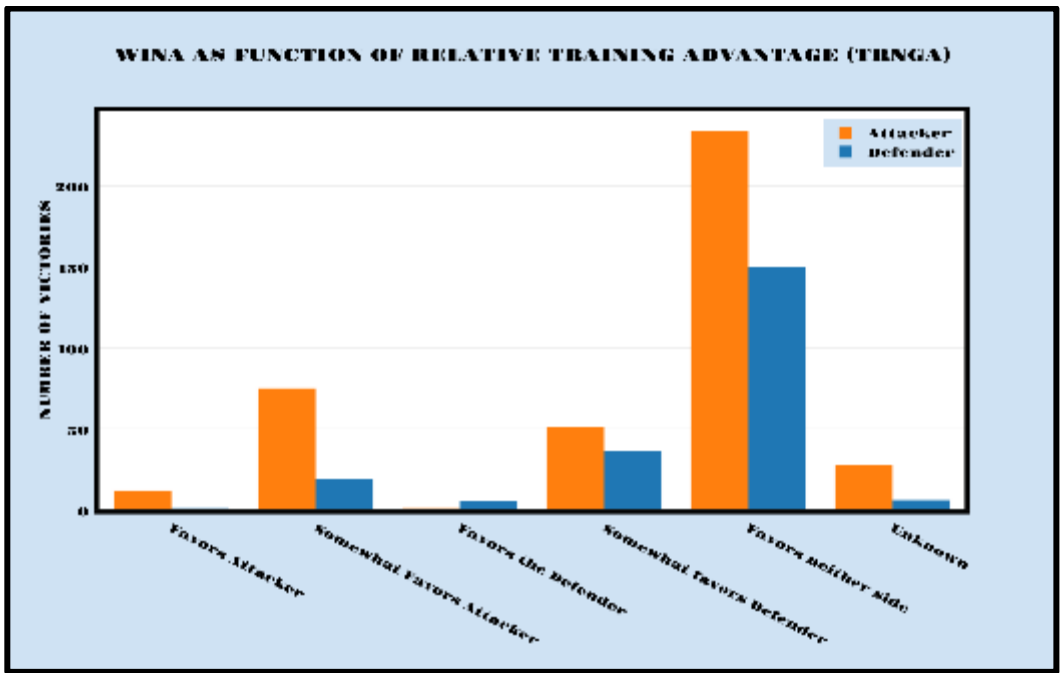
Table 32. Relative Training Advantage

RELATIVE TRAINING ADVANTAGE				
	Total	Attacker Wins	Defender Wins	Draw
Favors the attacker	13	12	1	0
Somewhat Favors the attacker	98	75	19	4
Favors the defender	6	1	5	0
Somewhat favors the defender	96	51	36	9
Favors neither side	413	234	150	29
Unknown	36	28	6	2

Figure 29 shows that most of the battles are fought without each side enjoying any relative training advantage. When the attacker holds the relative training advantage over the defender, results are more pronounced in favor of the attacker. In most of these

battles, the attacker was the winner. This is because defense does not require as much coordination as offense.

Figure 29. WINA as a Function of Relative Training Advantage



23. Technology Advantage

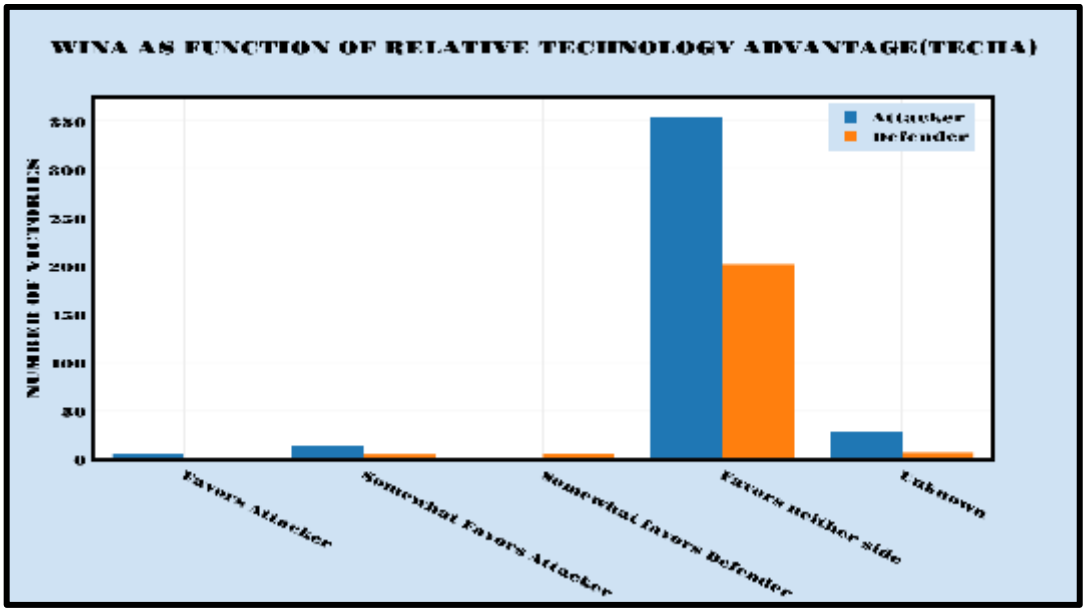
Technology stands for the development of modern weapons systems in a different time frame. Technology remains different in each century. During the seventeenth century, wars were mainly fought with musketeers and less precise artillery. However, in later centuries, automatics and more precise artillery guns replaced these weapons. History also witnessed the development of air power that resulted in third dimension warfare. In this dataset, most of the battles are fought with neither side having any advantage over the other. Whichever side holds the technological advantage is the side that has more victories (See Table 33).

Table 33. Relative Technology Advantage

ATTACKER RELATIVE TECHNOLOGY ADVANTAGE				
	Total	Attacker Wins	Defender Wins	Draw
Favors the attacker	6	5	1	0
Somewhat favors the attacker	20	13	4	3
Somewhat favors the defender	6	1	5	0
Favors neither side	594	354	201	39
Unknown	36	28	6	2

In most cases of conventional warfare, nations with similar capabilities fight with each other. There are rare occasions that nations with a large difference in technology are pitched against each other. Most of the battles are fought without either side enjoying any technological advantage (see Figure 30). In all situations, attackers win more than defenders.

Figure 30. WINA as a Function of Technology Advantage



24. Terrain

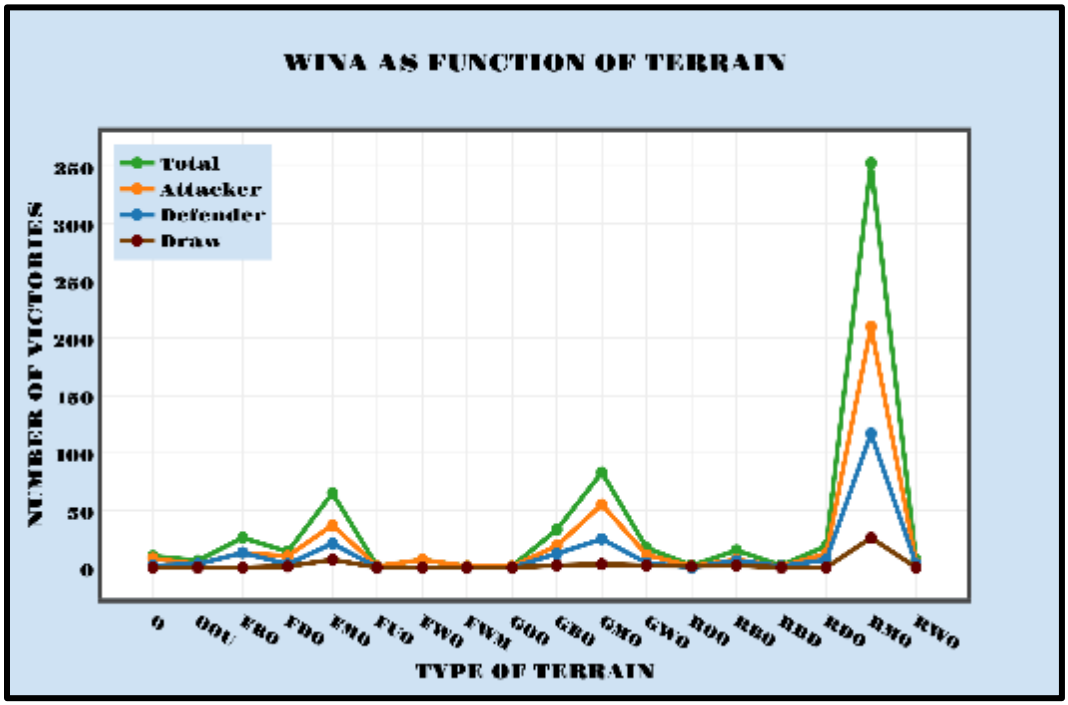
Great military leaders study terrain in detail; knowing the terrain is an obvious advantage for planning military campaigns. The planning of operations depends largely on the terrain. The type of formation, force composition, weapons, and technique for attack may all change depending on the type of terrain. An attack in mountains is different from an attack in plains or desert. There are 18 types of terrain in the dataset made up of three different characteristics (see Table 34). Most battles were fought in the RMO type (Rolling, Mixed, Not Available). There are 353 battles in this category, out of which attackers won 210, defenders won 117, and draws were 26.

Table 34. Terrain Description

TERRAIN DESCRIPTION							
	Character			Total	Attacker Wins	Defender Wins	Draw
	First	Second	Third				
OOO	NA	NA	NA	10	8	2	0
OOU	NA	NA	Urban	6	3	3	0
FB0	Flat	Bare	NA	26	13	13	0
FD0	Flat	Desert	NA	14	10	3	1
FM0	Flat	Mixed	NA	65	37	21	7
FU0	Flat	Urban	NA	1	1	0	0
FW0	Flat	Heavily Wooded	NA	7	7	0	0
FWM	Flat	Heavily Wooded	Mixed	1	1	0	0
G00	Rugged	NA	NA	1	1	0	0
GB0	Rugged	Bare	NA	33	19	12	2
GM0	Rugged	Mixed	NA	83	55	25	3
GW0	Rugged	Heavily Wooded	NA	17	11	4	2
R00	Rolling	NA	NA	2	1	0	1
RB0	Rolling	Bare	NA	15	7	6	2
RBD	Rolling	Bare	Desert	2	1	1	0
RD0	Rolling	Desert	NA	19	12	7	0
RM0	Rolling	Mixed	NA	353	210	117	26
RW0	Rolling	Heavily Wooded	NA	7	4	3	0

Figure 31 shows that most of the battles are within three levels: RM0 (Rolling, Mixed, Not Available), GM0 (Rugged, Mixed, Not Available), and FM0 (Flat, Heavily Wooded, Not Available).

Figure 31. WINA as a Function of Terrain



25. Weather

The weather variable has 49 levels that comprise five different characters (see Table 35). The first character is dry or wet weather conditions. The second character deals with the specifics of precipitation. The third character is the local temperature. The fourth character is for general climate, which includes spring, summer, and winter. The fifth character is another general climate factor. Most of battles (145) were fought in DSTST (Dry, Sunny, Temperate, Summer, Temperate). In these battles, the attacker won 85, the defender won 52, and there were eight draws.

Table 35. Weather Description

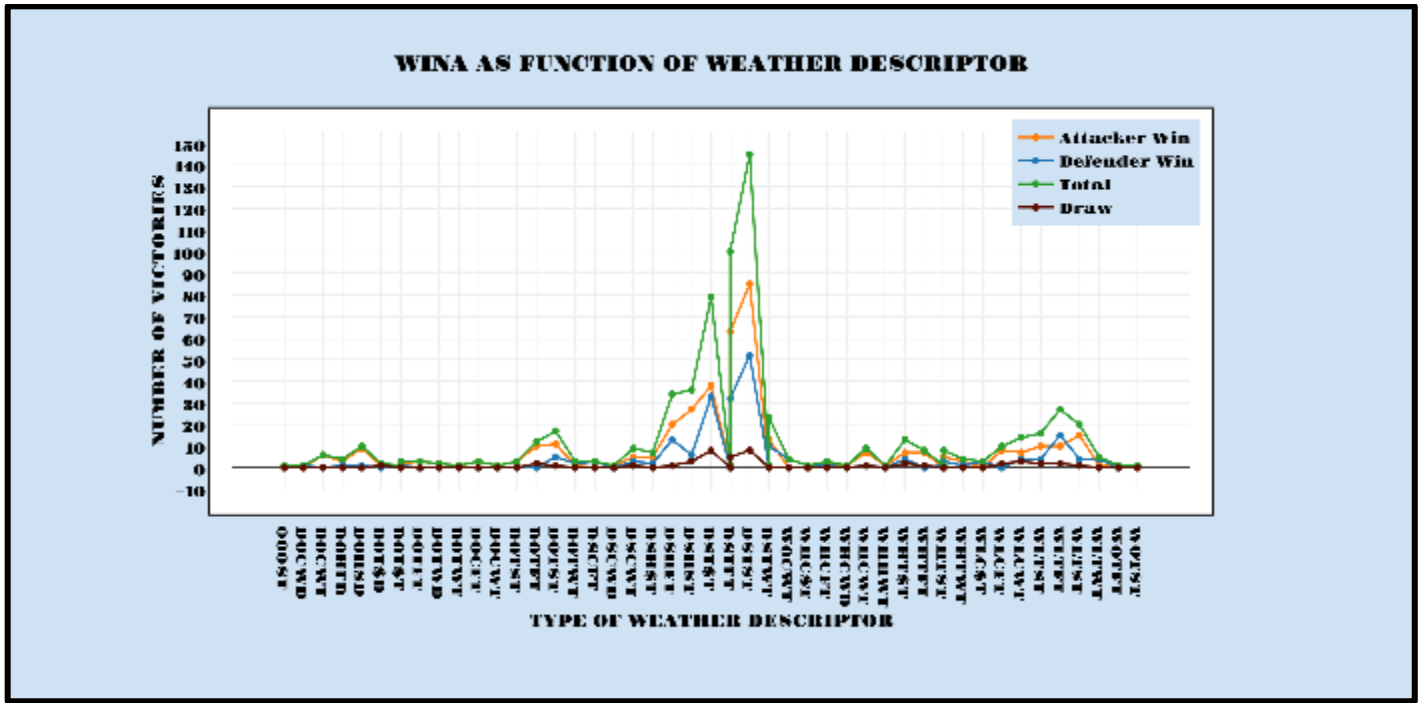
WEATHER DESCRIPTION									
	Character					Total	Attacker Wins	Defender Wins	Draw
	First	Second	Third	Fourth	Fifth				
OOOST	NA	NA	NA	Summer	Temperate	1	0	1	0
DOCWD	Dry	NA	Cold	Winter	Desert	1	0	1	0
DOCWT	Dry	NA	Cold	Winter	Temperate	6	6	0	0
DOHFD	Dry	NA	Hot	Fall	Desert	4	3	1	0
DOHSD	Dry	NA	Hot	Summer	Desert	10	9	1	0
DOT$D	Dry	NA	Temperate	Spring	Desert	2	1	0	1
DOT$T	Dry	NA	Temperate	Spring	Tropical	1	1	0	0
DOT$T	Dry	NA	Temperate	Spring	Temperate	3	2	1	0
DOTFT	Dry	NA	Temperate	Fall	Temperate	3	3	0	0
DOTWD	Dry	NA	Temperate	Winter	Desert	2	2	0	0
DOTWT	Dry	NA	Temperate	Winter	Temperate	1	1	0	0
DOCFT	Dry	Overcast	Cold	Fall	Temperate	3	3	0	0
DOCWT	Dry	Overcast	Cold	Winter	Temperate	1	1	0	0

	Character					Total	Attacker Wins	Defender Wins	Draw
	First	Second	Third	Fourth	Fifth				
DOT$T	Dry	Overcast	Temperate	Spring	Temperate	3	3	0	0
DOTFT	Dry	Overcast	Temperate	Fall	Temperate	12	10	0	2
DOTST	Dry	Overcast	Temperate	Summer	Temperate	17	11	5	1
DOTWT	Dry	Overcast	Temperate	Winter	Temperate	3	1	2	0
DSCFT	Dry	Sunny	Cold	Fall	Temperate	3	0	3	0
DSCWD	Dry	Sunny	Cold	Winter	Desert	1	1	0	0
DSCWT	Dry	Sunny	Cold	Winter	Temperate	9	5	3	1
DSH$T	Dry	Sunny	Hot	Spring	Temperate	7	5	2	0
DSHFT	Dry	Sunny	Hot	Fall	Temperate	34	20	13	1
DSHST	Dry	Sunny	Hot	Summer	Temperate	36	27	6	3
DST$T	Dry	Sunny	Temperate	Spring	Temperate	79	38	33	8
DSTFT	Dry	Sunny	Temperate	Fall	Tropical	1	1	0	0
DSTFT	Dry	Sunny	Temperate	Fall	Temperate	100	63	32	5
DSTST	Dry	Sunny	Temperate	Summer	Temperate	145	85	52	8
DSTWT	Dry	Sunny	Temperate	Winter	Tropical	1	1	0	0
DSTWT	Dry	Sunny	Temperate	Winter	Temperate	23	13	10	0
W0CWT	Wet	NA	Cold	Winter	Temperate	4	0	4	0
WHC$T	Wet	Heavy Precipitation	Cold	Spring	Temperate	1	0	1	0
WHCFT	Wet	Heavy Precipitation	Cold	Fall	Temperate	3	2	1	0
WHCWD	Wet	Heavy Precipitation	Cold	Winter	Desert	1	1	0	0
WHCWT	Wet	Heavy Precipitation	Cold	Winter	Temperate	9	7	1	1
WHHWT	Wet	Heavy Precipitation	Hot	Winter	Tropical	1	1	0	0
WHT$T	Wet	Heavy Precipitation	Temperate	Spring	Temperate	13	7	4	2
WHTFT	Wet	Heavy Precipitation	Temperate	Fall	Temperate	8	7	0	1

	Character					Total	Attacker Wins	Defender Wins	Draw
	First	Second	Third	Fourth	Fifth				
WHTST	Wet	Heavy Precipitation	Temperate	Summer	Tropical	1	0	1	0
WHTST	Wet	Heavy Precipitation	Temperate	Summer	Temperate	8	5	3	0
WHTWT	Wet	Heavy Precipitation	Temperate	Winter	Temperate	4	3	1	0
WLC$T	Wet	Light Precipitation	Cold	Spring	Temperate	3	0	3	0
WLCFT	Wet	Light Precipitation	Cold	Fall	Temperate	10	8	0	2
WLCWT	Wet	Light Precipitation	Cold	Winter	Temperate	14	7	4	3
WLT$T	Wet	Light Precipitation	Temperate	Spring	Temperate	16	10	4	2
WLTFT	Wet	Light Precipitation	Temperate	Fall	Temperate	27	10	15	2
WLTST	Wet	Light Precipitation	Temperate	Summer	Temperate	20	15	4	1
WLTWT	Wet	Light Precipitation	Temperate	Winter	Temperate	5	1	4	0
WOTFT	Wet	Overcast	Temperate	Fall	Temperate	1	1	0	0
WOTST	Wet	Overcast	Temperate	Summer	Temperate	1	0	1	0

Figure 32 is the visualization of Table 35, showing battles in different weather conditions.

Figure 32. WINA as a Function of Weather Descriptor



THIS PAGE INTENTIONALLY LEFT BLANK

IV. SUMMARY STATISTICAL ANALYSIS OF IRREGULAR WARFARE DATASETS

Irregular warfare is different in many aspects from regular warfare. These differences lead to different analytical approaches for driving quantifiable results from the datasets. In order to identify the important variables influencing the outcome of irregular warfare, it should be studied in depth from different perspectives and approaches. In this chapter, different datasets are analyzed to develop an understanding of irregular warfare. These datasets are unique and were compiled by different organizations. By analyzing different datasets, we are able to form a common picture of irregular conflicts. The recording of data in conflicts is a difficult task, and it becomes more daunting when the other side involved is a non-state actor. In the case of conventional war, data can be compiled using archives and official records of either side. In the case of irregular warfare, access to non-state actor data is limited. In simple terms, data collection for irregular warfare is difficult and prone to more mistakes than regular warfare data.

A. STATISTICAL ANALYSIS OF ARMED CONFLICT AND EVENT DATA PROJECT (ACLED)

Armed Conflict and Event Data (ACLED) is a collection of data related to conflicts in developing countries.[93] ACLED primarily focuses on conflict and political violence in African countries from 1997 to present day. The project also has information on some Asian countries. ACLED is the combined effort of many analysts under the leadership of Professor Clionadh Raleigh at the University of Sussex.[94] The data are derived from a variety of different sources, which include media, humanitarian agencies, and research publications. The project presently has five different variants of its dataset. Versions one and five of the dataset contain information about conflicts in Afghanistan, Cambodia, Haiti, Laos, Myanmar, Nepal, Pakistan, and all African countries. This

[93] Clionadh Raleigh, Linke Andrew, Hegre Havard and Karlsen Joakim, "Armed Conflict Location and Event Data," Journal of Peace Research 47(5) 1–102015.

[94] Ibid.

chapter focuses on versions one and five of the dataset. ACLED is human-coded and contains conflict event data related to various groups involved in conflicts; violent activities between non-state groups, records of attacks on civilians, records of territorial control between state and non-state agents; establishment of military bases; and information pertaining to riots and protests. The actors and events mentioned in the datasets of ACLED are elaborately explained in its codebook. However, for the reader's convenience, each variable described in the ACLED codebook is briefly discussed in the succeeding paragraphs.[95]

1. Actors

ACLED distinguishes between actors including the government, rebels, militia, active political organizations, and civilians.[96] These actors interact with each other over various issues seeking power. The governments in the dataset are recognized as sovereign and legitimate regimes. The rebels in the dataset are defined as groups of individuals acting against sovereign governments. Rebels use violence as their primary means and are recognized by people other than their own group. The specific groups use political militias for violence.[97] Militias do not normally seek the dismissal of governments; rather, they are instruments of power in the hands of the elite for political movements.[98] The dataset also contains information about unidentified armed groups, identity militias, civilians, and others (e.g., hired mercenaries and private security firms).

2. Events

An event, as defined by ACLED, "is the use of force by one or more groups for a political end although there are some instances, like protests, and non-violent activities, that are also included in the dataset."[99] An event occurs between two or more actors at a

---

95 Clionadh Raleigh and Caitriona Dowd, "Armed Conflict and Event Data Project (ACLED) Codebook" (2015).

96 Ibid.

97 Ibid.

98 Ibid.

99 Ibid.

specific location and time. ACLED has divided these events into the following nine sub-events:

a. Battle No change of Territory[100]
b. Battle Non-State Actors Overtake Territory[101]
c. Battle Government Regains Territory[102]
d. Headquarters or Base Established[103]
e. Non-Violent Activity by Conflict Actor[104]
f. Riots /Protests[105]
g. Violence Against Civilians[106]
h. Non-Violent Transfer of Territory[107]
i. Remote Violence[108]

B. STATISTICAL ANALYSIS OF DATA FOR AFGHANISTAN

Afghanistan is a landlocked country located in South Asia, west of Pakistan and east of Iran. Afghanistan has been embroiled in political violence and conflict at least since 1979, after its invasion by the Soviet Union.[109] The Soviet defeat and withdrawal did not return peace to Afghanistan. The period after the withdrawal of the Soviet Union witnessed political instability and civil war resulting in the emergence of the Taliban government. The U.S. and International Security Assistance Force (ISAF) overthrew the Taliban government following the September 11 attacks. The United States and its partners remained in Afghanistan since 2001 and have faced a mounting insurgency from

100 Ibid.
101 Ibid.
102 Ibid.
103 Ibid.
104 Ibid.
105 Ibid.
106 Ibid.
107 Ibid.
108 Ibid.
109 Joseph J. Collins, *The Soviet Invasion of Afghanistan: A Study in the use of Force in Soviet Foreign Policy*, 1986).

various actors. The insurgency has resulted in heavy losses to troops from the international coalition as well as to the civilian population.

The dataset contains 17 variables and information from 2004 to 2010 about insurgents and politically violent activities in Afghanistan. There is no "win or lose" variable in the dataset regarding the outcome of certain events. This limitation restricts the analysis to a more confined statistical exploration of data and identification of trends.

1. Type of Events in Afghanistan

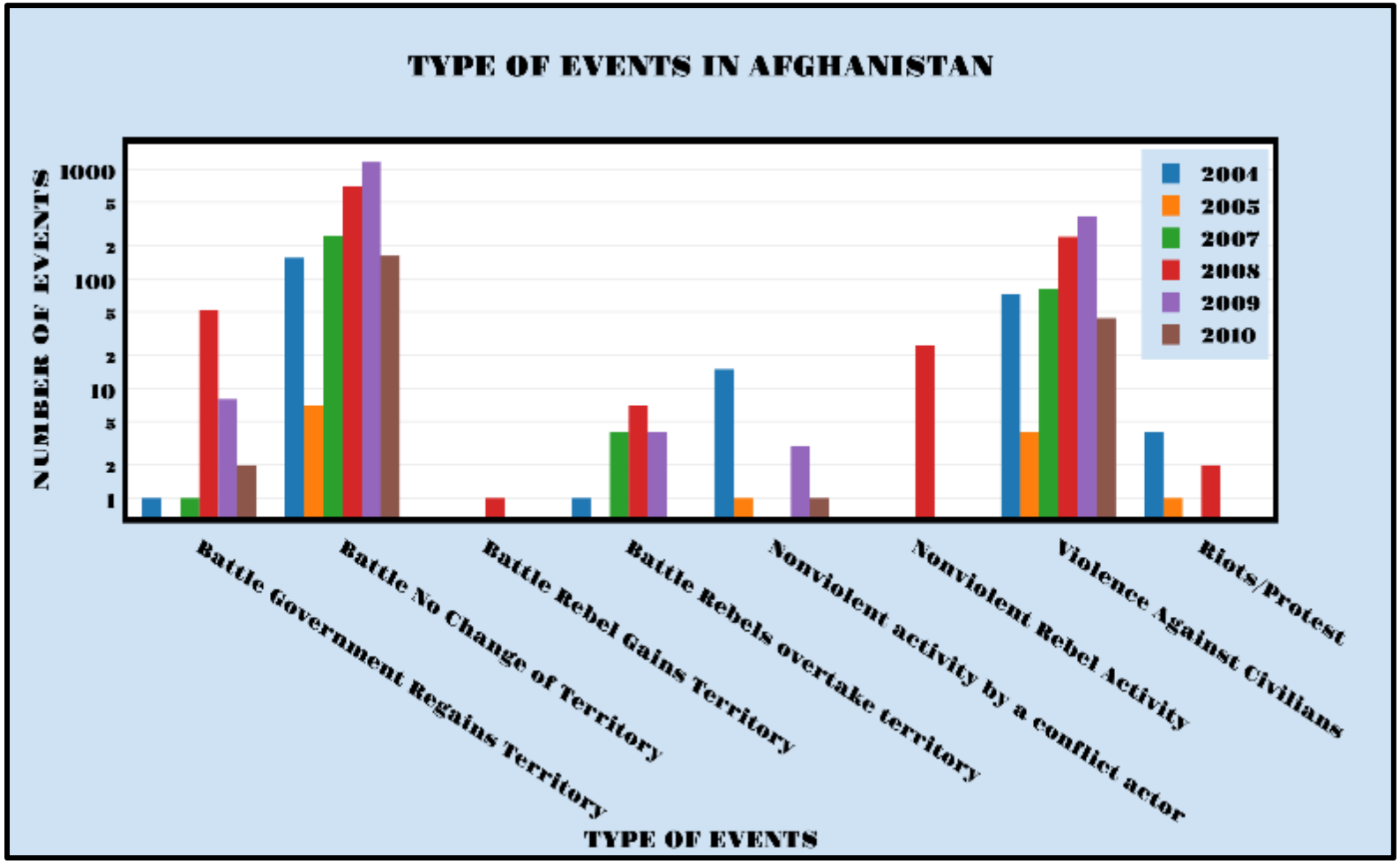
The dataset for Afghanistan contains eight different categories of events. The detailed description of these events has already been given in the description of the ACLED dataset. The maximum number of events occurred in 2009, and the second most in 2008. The total dataset contains 820 events of violence against civilians, and out of these, 618 took place from 2008 to 2009 alone. During this time frame, there was a majority of *battles no change of territory*. One can infer from this that the insurgents tried to subjugate security forces but failed. This led to a breakdown in the strength of the insurgent capabilities.

Table 36. Type of Events in Afghanistan

TYPE OF EVENTS IN AFGHANISTAN							
Type of Event	2004	2005	2007	2008	2009	2010	Total
Battle-Government Regains Territory	1		1	52	8	2	64
Battle-No Change of Territory	156	7	247	699	1159	163	2431
Battle-Rebel Gains Territory				1			1
Battle-Rebels overtake territory	1		4	7	4		16
Non–violent activity by a conflict actor	15	1			3	1	20
Non-Violent Rebel Activity				25			25
Violence Against Civilians	73	4	81	243	375	44	820
Riots/Protest	4	1		2			7
Total	250	13	333	1029	1549	210	3384

Figure 33 indicates different events in Afghanistan from 2004 to 2010. The most *violence against civilians* is done in the year 2009. The most number of incidents of *taking over territory by rebels* took place in 2008. The population has always been considered the focal point of insurgency, and violence against the population is normally carried out to undermine the legitimacy of the government. In the case of Afghanistan, the struggle of the insurgents was against the government, which had foreign support. As a result, violence was used as an instrument to coerce the population and undermine the authority of the government. Another important conclusion that can be drawn from this is that in irregular warfare, force attrition can be attributed to civilian protection missions. There is a consumption of force in protecting civilians.

Figure 33. Type of Events in Afghanistan between 2004 and 2010



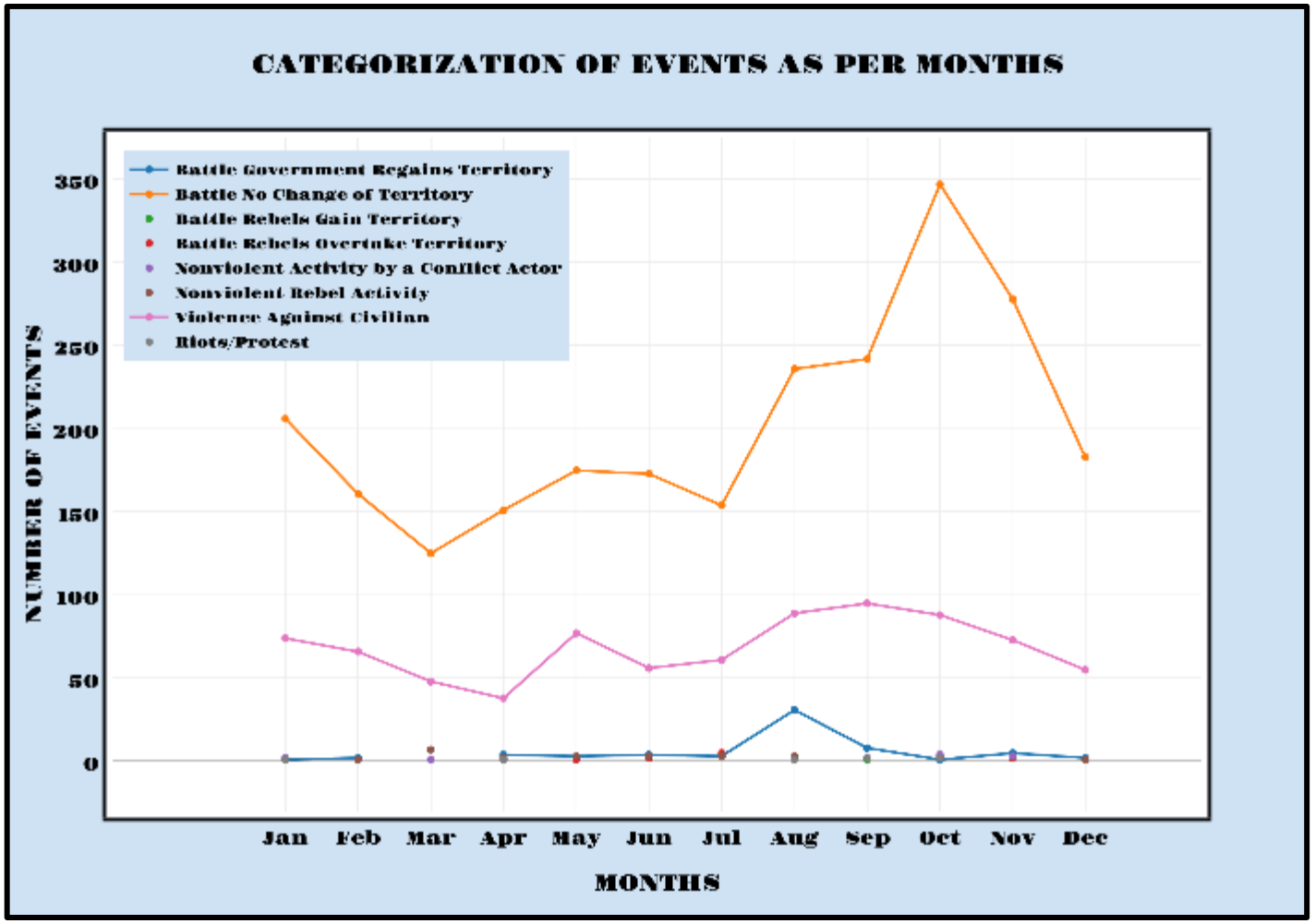
2. Seasonality Effect in Afghanistan

Afghanistan has harsh terrain and weather. For these reasons, forces normally carry out operations in summer. In order to validate this hypothesis, data were analyzed for a seasonality effect. For detailed results, see Table 37 and Figure 34. From the results, we can see that the peak violence against civilians occurs between August and October, after which it decreases. The maximum attempts by rebels to overtake territory fall between May and August. The most battles with *no change in territory* occur in October, after which there is a downfall in occurrence. Figure 34 indicates that December to March is the season of relative inactivity.

Table 37. Categorization of Events per Month

CATEGORIZATION OF EVENT PER MONTH									
Months	Battle Government Regains Territory	Battle No Change of Territory	Battle Rebel Gains Territory	Battle Rebels overtake territory	Nonviolent activity by a conflict actor	Nonviolent Rebel Activity	Violence Against Civilians	Riots/ Protest	Total
Jan	1	206			2	1	74	1	285
Feb	2	161			1	1	66		231
Mar		125			1	7	48		181
Apr	4	151			1	2	38	1	197
May	3	175		1		3	77		259
Jun	4	173		2	3	3	56		241
Jul	3	154		5		3	61		226
Aug	31	236		3	2	3	89	1	365
Sep	8	242	1		2		95	2	350
Oct	1	347		3	4		88	2	445
Nov	5	278		2	3		73		361
Dec	2	183			1	1	55		242

Figure 34. Categorization of Events as per Month



3. Actors Involved in Conflict

Irregular warfare differs from conventional in terms of the actors involved. Conventional warfare is normally between two countries or between countries against other groups of countries. However, in irregular warfare, there are numerous actors that play significant roles. The Afghan situation is a quagmire that illustrates how complex irregular warfare is due to the number of different players involved in it. One of the most important actors that appears from the analysis of the Afghanistan data is the civilian population, because in most of the actions civilians are target.

The ACLED dataset does not significantly differentiate between two sides as in contrast to the CDB90G dataset in which the attacker and defender were clearly identified. The dataset contains multiple entries for the ISAF partners that have been condensed to ISAF, and the same methodology is used for other key players. The data also contain multiple entries for the Afghan government based on the time period, which are also condensed to the Afghan Government. The analysis of both sides has been carried out and a relationship matrix has been established to identify the key players in Afghanistan's conflict (see Table 38).

The Taliban (Afghanistan) is the largest player, with a total of 1004 conflicts with various actors. The Taliban has 341 conflicts with civilians, 229 with police forces, 193 with ISAF, 143 with Afghanistan armed forces, 32 with the Afghan Government, 28 with unknown forces, 27 with private security, five with the United Nation Mission in Afghanistan (UNMA), two with an unidentified armed group of Afghanistan, two with tribes, one with the Central Intelligence Agency (CIA), and one with Afghan local militia. After the Taliban, the largest major player is unidentified armed groups of Afghanistan.

Table 38. Relationship Matrix Afghan Actors

RELATIONSHIP MATRIX AFGHAN ACTORS																								
	Afghan Government	**Afghan Local Militia**	**Afghanistan Armed Forces**	**Afghan Tribe**	**Al-Qaeda**	**International Security Assistance Force (ISAF)**	**Chechen Militia**	**CIA: Central Intelligence Agency (USA)**	**Civilians**	**Foreign Fighters**	**Foreign Journalists**	**Haqqani Network**	**Hizb-e-Islami**	**ICRC: International Committee of the Red Cross**	**Islamic Movement of Uzbekistan**	**Police Forces of Afghanistan**	**Private Security**	**Rioters (Afghanistan)**	**Security Firm (United States)**	**Taliban (Afghanistan)**	**UNAMA: United Nations Assistance Mission in Afghanistan**	**Unidentified Armed Group (Afghanistan)**	**Unknown**	Total
Afghan Government		1																		1				2
Afghan Local Militia	3	3	1	3					2											1		1	1	15
Afghan Tribe		1							1															2
Afghanistan Armed Forces																				3				3
Al-Qaeda			1																				1	2
Armed Forces of									1											26		2		29

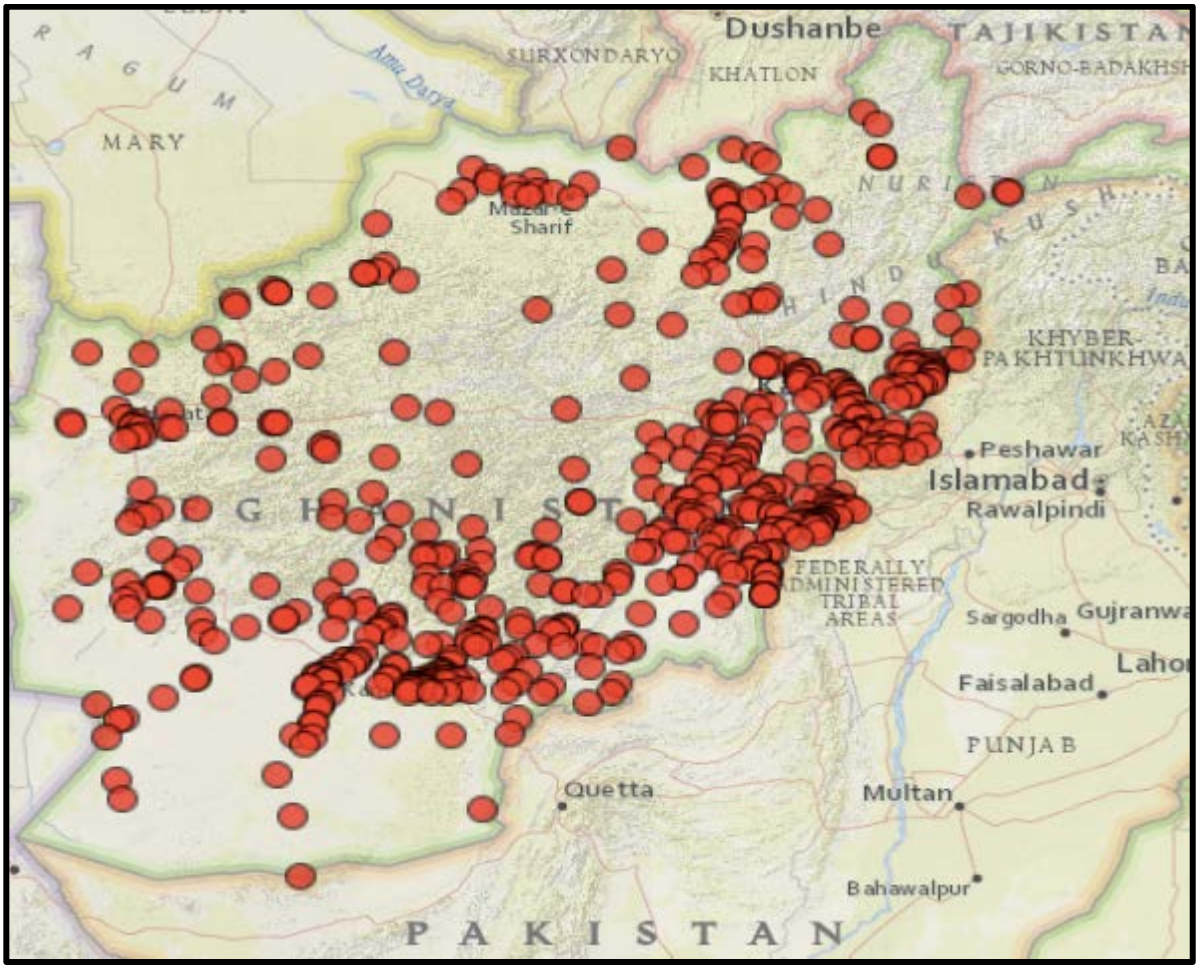
Afghanistan																								
International Security Assistance Force			1		2		1		60			8	3		1	3			1	47		33	1	589
Civilians (Afghanistan)		1	1													1				10		8	2	23
Foreign Fighters		1																						1
Hizb-e-Islami			1			8			2							3	1							15
International Security Assistance Forces			3	1					71	1		1	1			4				92		42		216
Military Forces of Afghanistan		2			1	3			13	3		1			2	1	1			17		10	3	306
National Department of Security of Afghanistan																						1		1
Police Forces of Afghanistan		2	3						3				1							11		35	1	160
Private Security																		1		1		1		3
Protesters (Afghanistan)			1																					1
Saad Abu Fourkan						1																		1
Taliban (Afghanistan)	3	1	14	2		19		1	34							22	2				5	2	2	100 4
Taliban (Pakistan)			1			2			1								1							5
Unidentified Armed Group (Afghanistan)	2	3	83	2	1	44			28		1			1		13	1			2	1	1	1	997
US Protection and																				2				2

Investigation																								
Unknow		1				1			2											3				7
Total	5	1	23	8	4	65	1	1	77	4	1	1	5	1	3	37	4	1	1	90	6	23	3	338 4

4. Geographical Location of Events in Afghanistan

The dataset contains additional information about locations of different events in Afghanistan. This information can be useful in determining the high, medium, and low threat areas. The geographical locations become more relevant in irregular warfare because of the significant role of the population in determining the outcome of conflicts as well as the lack of established front lines. In this case, most of the events were close to Pakistan. This may be due to the porous border and tribal affiliations between the two countries.

Figure 35. Geographical Location of Events in Afghanistan



C. STATISTICAL ANALYSIS OF DATA FOR CAMBODIA

Cambodia is located in Southeast Asia. Cambodia is bordered in the northwest by Thailand, northeast by Laos, east by Vietnam, and in the southwest by the Gulf of Thailand.[110] Elections were held in Cambodia in 1993 in accordance with the Paris Peace Accord of 1991. Prince Norodom Ranariddh's party, named the National United Front for an Independent, Neutral, Peaceful, and Cooperative Cambodia (FUNCINPEC), secured the most votes, followed by Hun Sen's Cambodian People's Party and the Buddhist Liberal Democratic Party.[111] After a brief period of political negotiation, Prince Ranariddh became the first, and Hun Sen the second, prime minister of Cambodia. This arrangement did not last long, and in 1997, fighting between their two political parties erupted. This resulted in a short but politically violent conflict between various players in Cambodian politics. This conflict was smaller than the 1970 conflict. However, it still resulted in injuries and lost lives.

The ACLED dataset for Cambodia contains 17 variables. The data are organized using the same pattern as the sets for Afghanistan and other countries. The dataset is explored statistically to visualize various trends such as the type of events, seasonality, role of different actors, and geographical locations.

1. Type of Events in Cambodia from 1997 to 2010

There are seven different kinds of events that happened in Cambodia from 1997 to 2010. The most common type of event was the *battle no change of territory*, which took place 140 times. The second most common event was *violence against civilians*, which took place 111 times. The third most common event recorded in the database was *riots/protest*, which took place 86 times. For a detail summary of events, see Table 39.

110 D David P Chandler, *The Land and People of Cambodia* (New York, NY: HarperCollins Publishers, 1991).

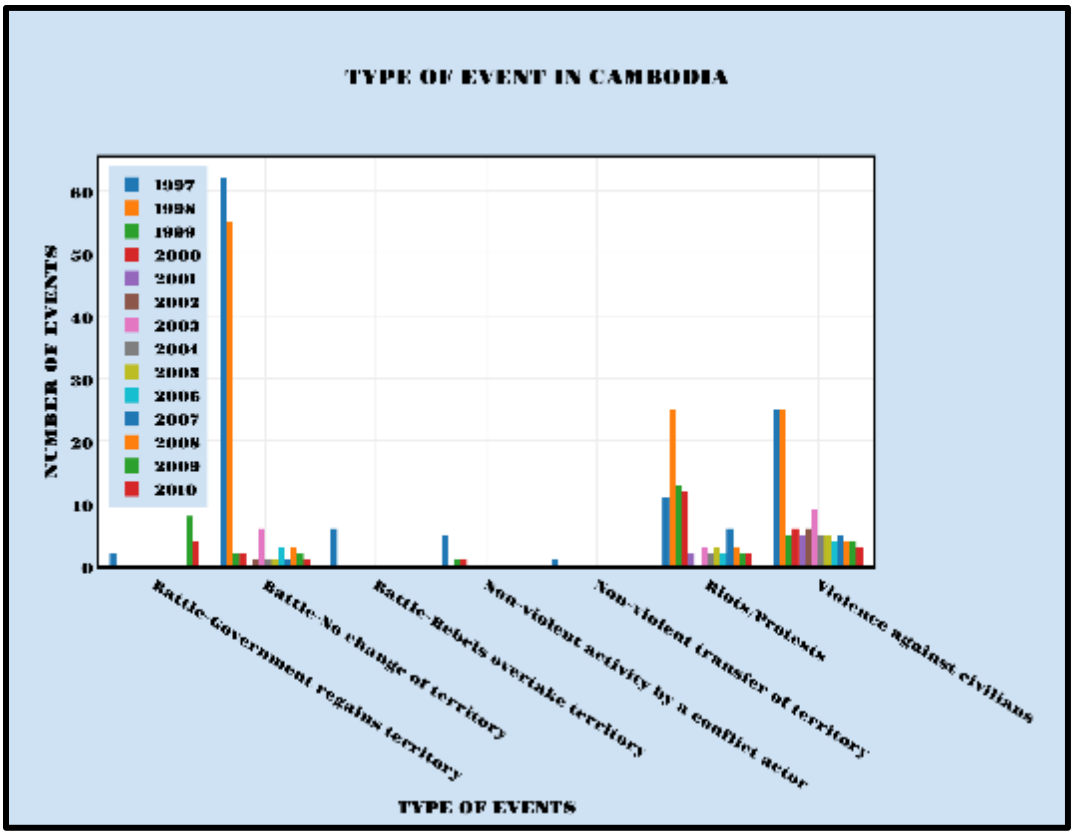
111 Kheang Un, *Democratization without Consolidation: The Case of Cambodia, 1993–2004*(DeKalb, IL: Northern Illinois University, 2004).

Table 39. Type of Events in Cambodia

TYPE OF EVENTS IN CAMBODIA															
Type of Event	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Total
Battle-Government regains territory	2												8	4	14
Battle-No change of territory	62	55	2	2		1	6	1	1	3	1	3	2	1	140
Battle-Rebels overtake territory	6														6
Non-violent activity by a conflict actor	5		1	1											7
Non-violent transfer of territory	1														1
Riots/ Protests	11	25	13	12	2		3	2	3	2	6	3	2	2	86
Violence against civilians	25	25	5	6	5	6	9	5	5	4	5	4	4	3	111
Total	112	105	21	21	7	7	18	8	9	9	12	10	8	10	357

Figure 36 displays trends for the different types of events. Most events were recorded between 1997 and 1998. The least violent years were between 2001 and 2002. Moreover, it can be concluded that there was a gradual decrease in the violent events in Cambodia over the span.

Figure 36. Type of Events in Cambodia between 1997 and 2010



2. Seasonality Effect in Cambodia

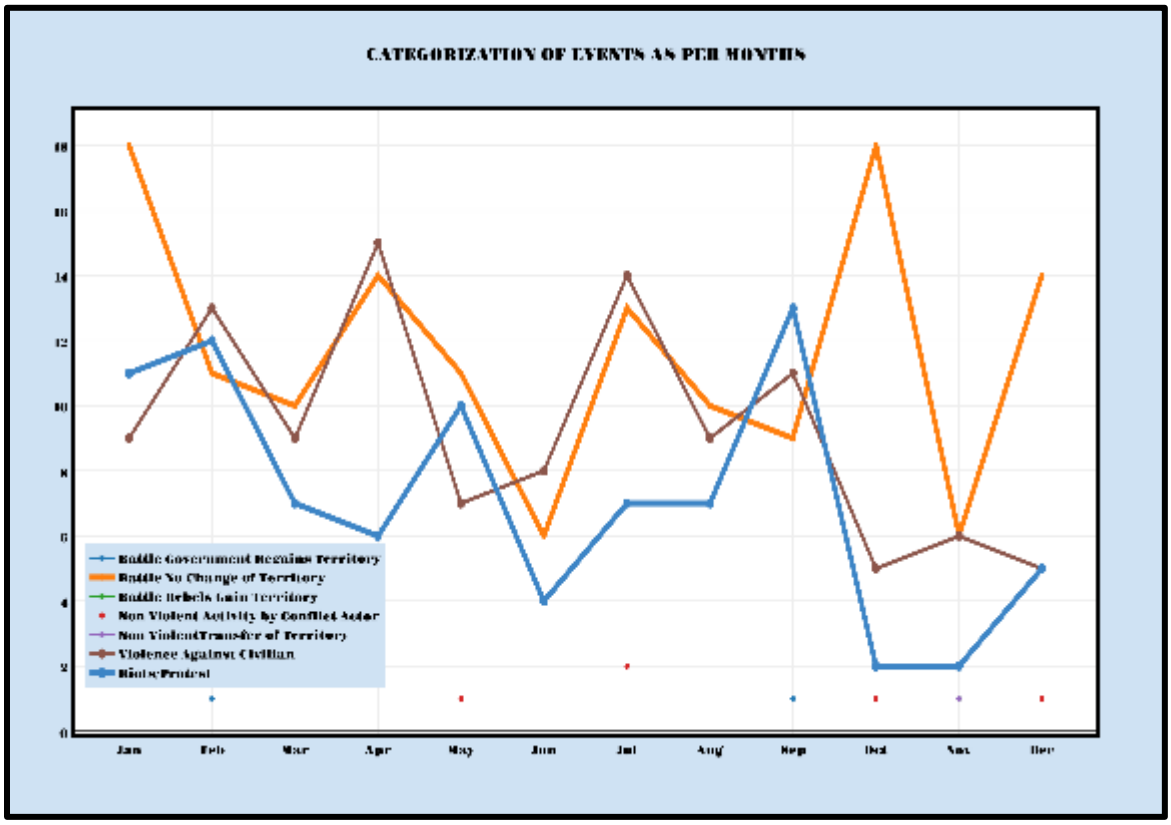
The most events took place in January and July. In both of these months, significant events were *battle no change of territory* and *violence against civilians*. However, most violence against civilians occurred in April. In April, there were 15 events of violence.

Table 40. Categorization of Events in Cambodia per Month

CATEGORIZATION OF EVENTS IN CAMBODIA PER MONTH								
Months	Battle –Government Regains Territory	Battle- No Change of Territory	Battle – Rebels overtake territory	Non –violent activity by a conflict actor	Non-Violent Transfer of Territory	Riots/Protest	Violence Against Civilians	Total
Jan		18				11	9	38
Feb	1	11				12	13	37
Mar		10				7	9	26
Apr	3	14		2		6	15	40
May		11		1		10	7	29
Jun		6				4	8	18
Jul		13	2	2		7	14	38
Aug		10				7	9	26
Sep	1	9	3			13	11	37
Oct		18		1		2	5	26
Nov		6			1	2	6	15
Dec	1	14	1	1		5	5	27
Total	6	140	6	7	1	86	111	357

A seasonality effect applies to Cambodia (see Figure 37). There are four visible peaks in the months of February, April, July, and October. Events can also be divided into small campaigns of violence between January to March, March to May, June to August, and September to December, which depict peaks in the middle of each bracket.

Figure 37. Categorization of Events in Cambodia per Month



3. Actors Involved in Conflict

Cambodia's relationship matrix is less complex than that of Afghanistan, but still involves a significant number of actors. One noticeable trend is that civilians are at the center of the conflict, with 91 events involving different actors. In 44 events, civilians encountered unidentified armed groups. Other important players in the game are FUNCINPEC, the Khmer Rogue Militia, and unknown forces.

Another noticeable trend in both Afghanistan and Cambodia datasets is the involvement of foreign actors in the conflicts. In Afghanistan, the Taliban of Pakistan was involved in carrying out attacks. In Cambodia, unidentified armed groups from Vietnam carried out attacks (see Table 41). We have already hypothesized that irregular warfare is a complicated phenomenon involving a large number of actors. This hypothesis seems to be validated by our initial examination of Afghanistan and Cambodia datasets.

Table 41. Relationship Matrix For Cambodian Actors

RELATIONSHIP MATRIX CAMBODIAN ACTORS																	
	Armed Tree Loggers (Cambodia)	Cambodian Freedom Fighters	Civilians (Cambodia)	CPP: Cambodian People's Party Militia	FUNCINPEC Royalist Militia	Khmer Rouge Militia	Military Forces of Cambodia	Military Forces of Thailand	Police Forces of Thailand	Protesters (Cambodia)	Protestors (Thailand)	Protestors (Vietnam)	Rioters (Cambodia)	SRP: Sam Rainsy Party (Cambodia)	Unidentified Armed Group (Cambodia)	Unknown	Total
Civilians		1		2	1	21	8	8	6						44		91
CPP: Cambodian People's Party			1												2		3
Democratic Front of Khmer Students and Intellectuals				1													1
FUNCINPEC: National United Front for an Independent, Neutral,				13			49								1		63

Peaceful, and Cooperative Cambodia																	
Khmer Rouge Militia			4	1		3	38										46
Khmer Serei Militia							1										1
KRAF: Khmer Royal Armed Forces						3										1	4
Military Forces of Cambodia	1	1	3			5		7							1		18
Military Forces of Thailand						1	8								1		10
Mutiny of Military Forces of Cambodia (1993–1997)							2										2
NUA: National United Army							4										4
Opposition Parties				1													1
Police Forces of Cambodia			1						1	22		1	2	1	1		29
Protesters (Cambodia)									2								2
Protestors (International)									1								1
Protestors (Vietnam)										1							1
SRP: Sam Rainsy Party (Cambodia)															2		2

Unidentified Armed Group (Cambodia)			11				1		1	1					1		15
Unidentified Armed Group (Vietnam)									1								1
Unknown				2	1	2	1		1	45	5		5				62
Total	1	2	20	20	2	35	112	15	13	69	5	1	7	1	53	1	357

4. Geographical Location of Events in Cambodia

There were 13 events that were not confined within the territorial limit of Cambodia. These events happened in Vietnam or close to the Vietnamese border (see Figure 38). Most events occurred close to the Vietnamese and Thai borders. This trend was also observed in Afghanistan, where most events were close to the Pakistani border. This can be interpreted as terrorists taking advantage of unguarded borders for movement and safe hide outs.

Figure 38. Geographical Location of Events in Cambodia

D. STATISTICAL ANALYSIS OF DATA FOR HAITI

Haiti is located in the Caribbean, and it occupies the western portion of the island of Hispaniola. Haiti shares Hispaniola with the Dominican Republic. In 1994, the U.S. government carried out Operation Uphold Democracy to restore the democratic government of President Jean-Bertrand Aristide.[112] The military operation under the United Nations (UN) mandate ended in March 1995 with a transfer of power to United Nations Mission in Haiti (UNMIH).

Elections were carried out in 1996 to restore democracy and transfer power to the people. However, things did not improve as the people had wished. In 1997, poverty, a low living standard, and economic conditions erupted in violence. Port-Au-Prince, the capital of Haiti, suffered violent protests, strikes, and gang war, which immediately spread all over the country. These conflicts led Haiti into another era of political violence and disturbance. The fight to achieve power among various groups resulted in irregular conflict.

1. Type of Events in Haiti

The Haiti dataset contains information between 1997 and 2010. It has 20 variables with a record of 1,076 different events. An analysis of the Haiti dataset was carried out in a similar way to Afghanistan and Cambodia. There are seven types of events in Haiti. Most of the events that took place in Haiti fall within the *riots/protests* category, followed by 344 events of *violence against civilians,* and 258 events of *no change in territory* (see Table 42).

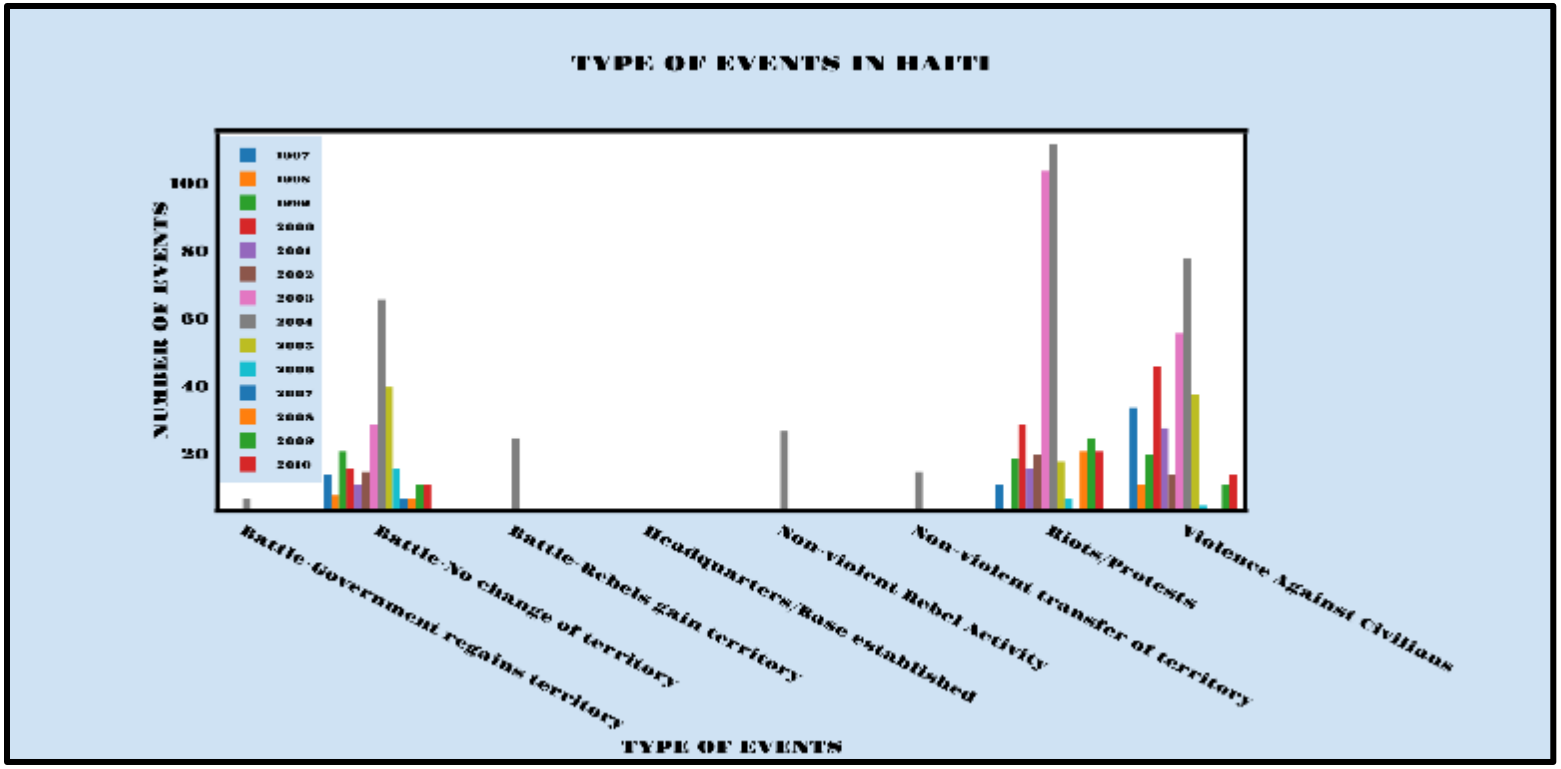
[112] Kretchik Edward Walter. *Invasion, Intervention," Intervasion": A Concise History of the U.S. Army in Operation Uphold Democracy* (Darby, PA: DIANE Publishing, 1998).

Table 42. Type of Events in Haiti

TYPE OF EVENTS HAITI															
Type of Event	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Total
Battle-Government regains territory								6	1						7
Battle-No change of territory	13	7	20	15	10	14	28	65	39	15	6	6	10	10	258
Battle-Rebels gain territory						1		24							25
Headquarters/ Base established						1									1
Non-violent Rebel Activity						1	2	26	2					1	32
Non-violent transfer of territory								14							14
Riots/Protests	10	2	18	28	15	19	103	111	17	6	2	20	24	20	395
Violence Against Civilians	33	10	19	45	27	13	55	77	37	4		1	10	13	344
Total	56	19	57	88	52	49	188	323	96	25	8	27	44	44	1076

Most events occurred in 2004. The second highest number of events took place in 2003. These two years are the peak of violence (see Figure 39). Despite these two years, the events remain fairly well below 100 incidents for the rest of the years.

Figure 39. Type of Events in Haiti between 1997 and 2010



2. Seasonality Effect in Haiti

Most events took place between December and February, with a record number of events within a month taking place in February (181 incidents). This trend corresponds with the coolest and driest month in Haiti, which is January.[113] The numbers of events between March and November are almost the same, except for those in July and August, which dip significantly. This decline correlates with weather because July and August are the hottest months in Haiti.

Table 43. Categorization of Events in Haiti per Month

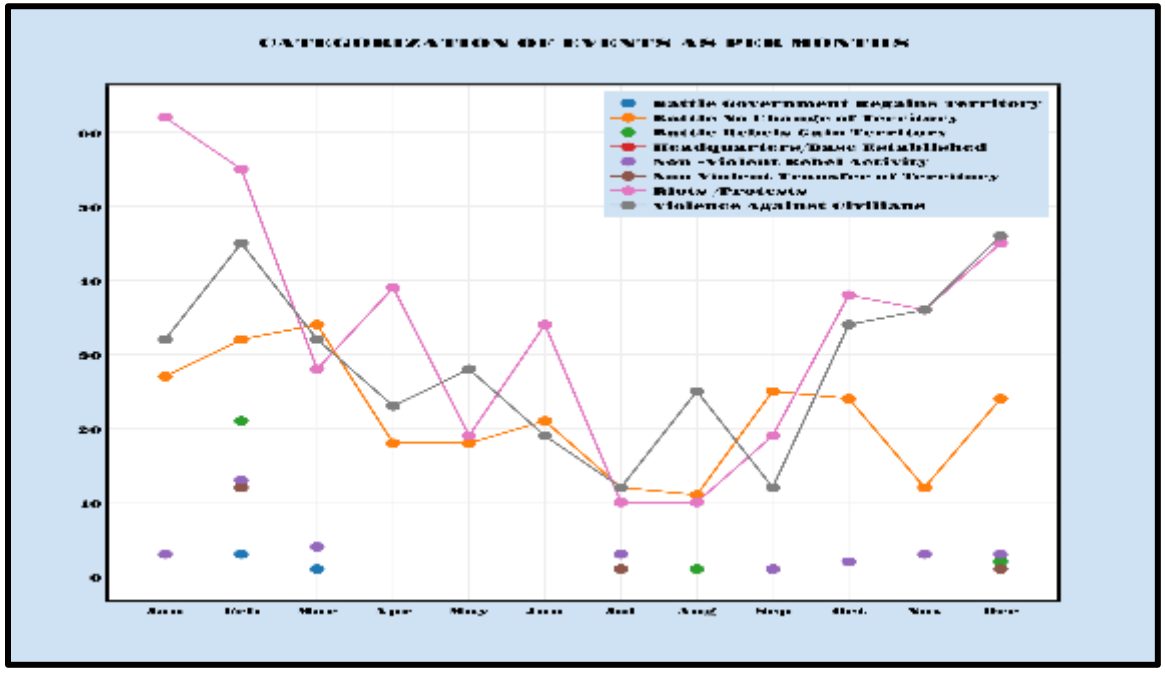
CATEGORIZATION OF EVENTS IN HAITI PER MONTH									
Months	Battle-Government regains territory	Battle-No change of territory	Battle-Rebels gain territory	Headquarters/Base established	Non-violent Rebel Activity	Non-violent transfer of territory	Riots/Protests	Violence Against Civilians	Total
Jan		27			3		62	32	124
Feb	3	32	21		13	12	55	45	181
Mar	1	34			4		28	32	99
Apr		18					39	23	80
May		18					19	28	65
Jun		21					34	19	74
Jul		12			3	1	10	12	38

113 "Port-Au-Prince, Haiti," World Weather and Climate Information., , http://www.weather-and-climate.com/average-monthly-Rainfall-Temperature-Sunshine,Port-Au-Prince,Haiti.

Aug		11	1				10	25	47
Sep	1	25	1	1	1		19	12	60
Oct		24			2		38	34	98
Nov		12			3		36	36	87
Dec	2	24	2		3	1	45	46	123
Total	7	258	25	1	32	14	395	344	1076

There is a gradual drop in violence against civilians in February, and from July to August, there is a flat trend in the events, after which the violence starts rising (see Figure 40).

Figure 40. Categorization of Events per Month



3. Actors Involved in Conflict

In Haiti's dataset, like Afghanistan's and Cambodia's, the population emerged as one of the major players. From the dataset, the most important actor in Haiti is the civilian population. Civilians have been in 287 different conflicts with other actors. After civilians, the second largest actor is the protester followed by the rioter (see Table 44). Rioters and protestors are documented as a separate group even though they are also civilians.

Table 44. Relationship Matrix Haiti Actors

RELATIONSHIP MATRIX HAITI ACTORS																									
	Aristide Loyalists Militia	Border Custom Enforcement in Haiti (1996–2001)	Cannibal Army militia	Christian Militia (Haiti)	Cite Soleil Militia (Haiti)	Civilians	Democratic Consultation Group (Espace) Militia	Democratic Convergence (Opposition Alliance) Militia	Front for Aristide's Departure (Cannibal Army)	Fwon Lespwa (Hope Front)	Lafanmi Se Lavi Militia	Lavalas Militia	Military Forces of Dominican Republic	Military Forces of the United States (2001–2009)	MINUSTAH: United Nations Stabilization Mission in	Mutiny of Police Forces of Haiti (1996–2001)	Police Forces of Dominican Republic	Police Forces of Haiti	Political Opposition Militia (Haiti)	Protesters (Haiti)	Rioters (Haiti)	Rioters (Lavalas)	Unidentified Armed Group (Haiti)	Unidentified Armed Group (Colombia)	Unidentified Armed Group (Dominican Republic)	Unidentified Political Militia (Haiti)	Vodouist Militia	Unknown

ACSM: Motherless Convergence Army												1						3											
Base Resistance																		3											
Cannibal Army militia																		4					1					1
Civilians				1	2			3	1		1	4	1		3		1	7	1		8		1		2	1		
Democratic Convergence (Opposition Alliance) Militia												1						3										
ESCANP/Korega: Effort and Solidarity To Build a National and Popular Alternative/ Grand'Anse Resistanc							1																					

e Committee Militia																												
Faction of Lavalas Militia (Dread Mackenzie)												1																
FLRN: National Liberation and Reconstruction Front						4									1			2					1					1
Former Military Forces of Haiti (Army)						3						1			7			9					2					4
Front for Aristide's Departure (Cannibal Army)																		1										
Gonaïves Resistance Front																		1										
Guy Philippe																		2										
Lavalas Militia						1		1						4	9			1										4
Military																							1					

Forces of Canada (2003–2006)																												
Military Forces of the Dominican Republic																						1						
Military Forces of The United States																				1								
MINUSTAH: United Nations Stabilization Mission in Haiti					1																	7						
Opposition Party Militia (Haiti)										1		3																
Police Forces of the Dominican Republic (2006-)		1				1		1	1			2				1				4	5	1	3					
Police Forces of the United																							1					

States (1994–2001)																												
Private Security Forces (Haiti)																					1							
Protesters (Haiti)																					1							
RAMICOS: Group of Principled Militants of St. Marc																	4											2
Revolutionary Artibonite Resistance Front												1					3											
Rioters (Dominican Republic)						1						1					1											
Rioters (Haiti)						1																						1
Rioters (Lavalas)						7						3			3		2					2	1	1				6
Unidentified Armed Group (Dominican						1																						

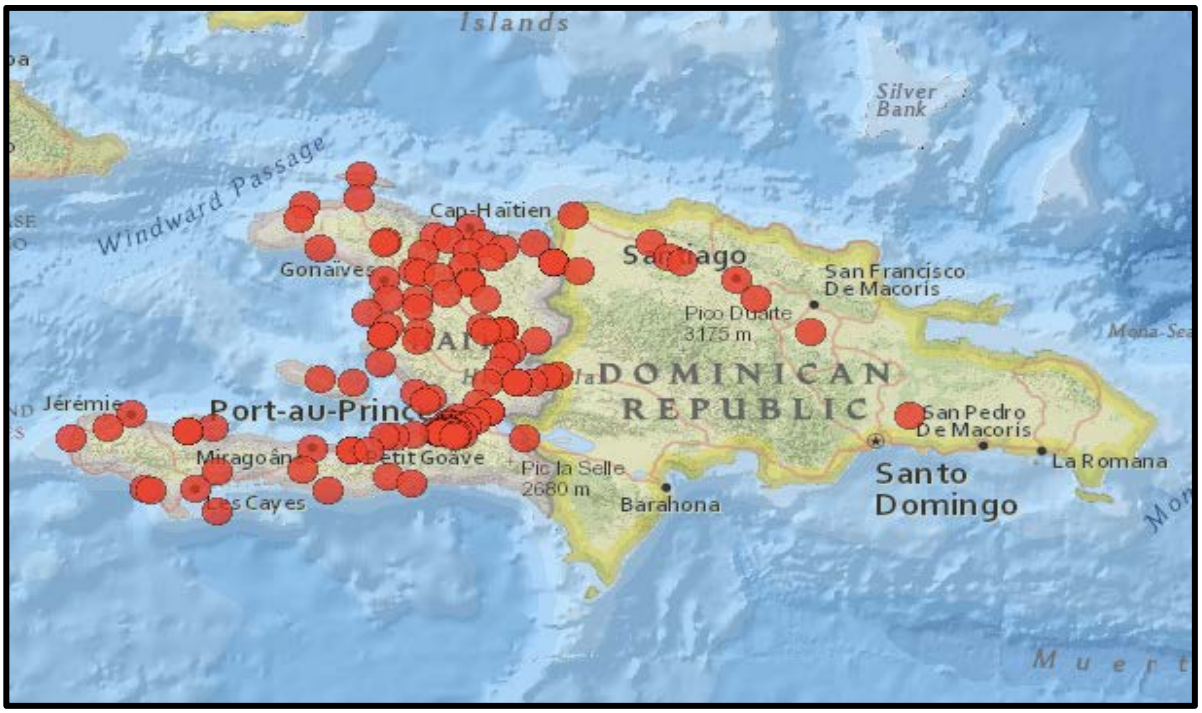
Republic)																												
Unidentified Armed Group (Haiti)						2															1							
Unidentified Armed Group II (Haiti)	1		1			1						2	1	3	1			4			2							7
Unidentified Political Militia (Haiti)																							1					
Unknown						1																						

Another important trend in the Haiti dataset, which is common also in Afghanistan and Cambodia, is the presence of foreign players in conflict. In the case of Haiti, we can see the presence of the Armed Group of the Dominican Republic.

4. Geographical Location of Events in Haiti

There are 14 events in the Haiti dataset that took place outside Haiti. This trend has been observed in other countries, including Afghanistan and Cambodia. From this trend it can be seen that in irregular warfare, sometimes incidents occur outside of the geographical limits of the primary country. Moreover, low, medium, and high conflict areas can be clearly identified.

Figure 41. Geographical Location of Events in Haiti



E. STATISTICAL ANALYSIS OF DATA FOR LAOS

Laos is a landlocked country bordering China to the northwest, Thailand to the west, Vietnam to the east, and Cambodia to the south. After independence from Japan in 1945, Laos went under French rule until 1949.[114] After achieving independence from France, Laos became a constitutional monarchy under Sisavang Vong. This government faced a civil war initiated by the communist party, which resulted in a regime change. Since 1975, Laos has been under communist rule. The dataset for Laos is a collection of politically violent events between 1998 and 2010.

1. Types of Events in Laos from 1998 to 2010

Laos's dataset is relatively small compared to the other datasets. There are only three types of events and 67 in total. However, like the other datasets, most events involved *violence against civilians*.

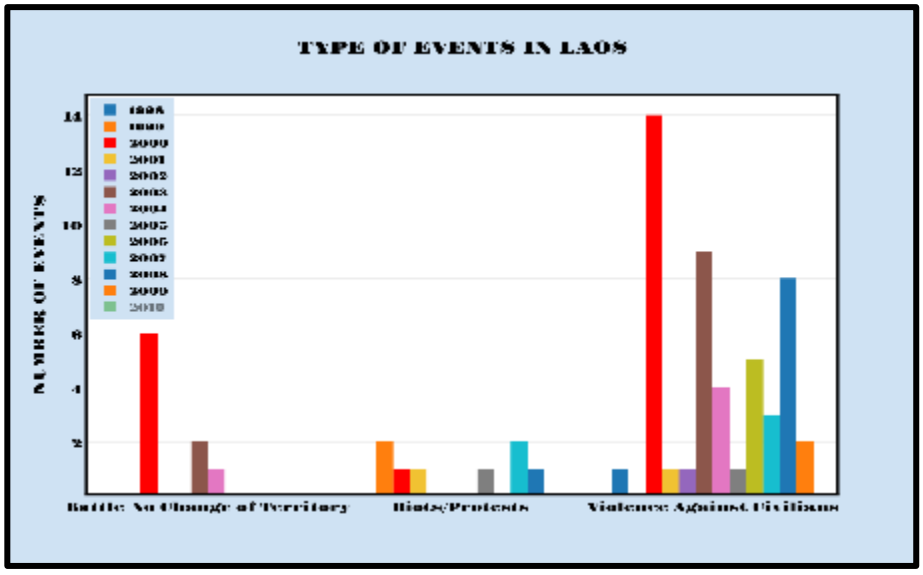
[114] Gunn, C. Geoffrey *Political struggles in Laos, 1930–1954: Vietnamese Communist Power and the Lao Struggle for National Independence* (Bangkok, Thailand: Editions Duang Kamol, 1988).

Table 45. Type of Events in Laos

TYPE OF EVENTS LAOS														
Type of Event	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Total
Battle-No Change of Territory			6			2	1							9
Riots/Protests		2	1	1				1		2	1			8
Violence Against Civilians	1		14	1	1	9	4	1	5	3	8	2	1	50
Total	1	2	21	2	1	11	5	2	5	5	9	2	1	67

Most of events took place in 2000. After 2000, the second most events took place in 2003. In other years the number of events are less than nine. (see Figure 42).

Figure 42. Types of Events in Laos between 1998 and 2010



2. Seasonality Effect in Laos

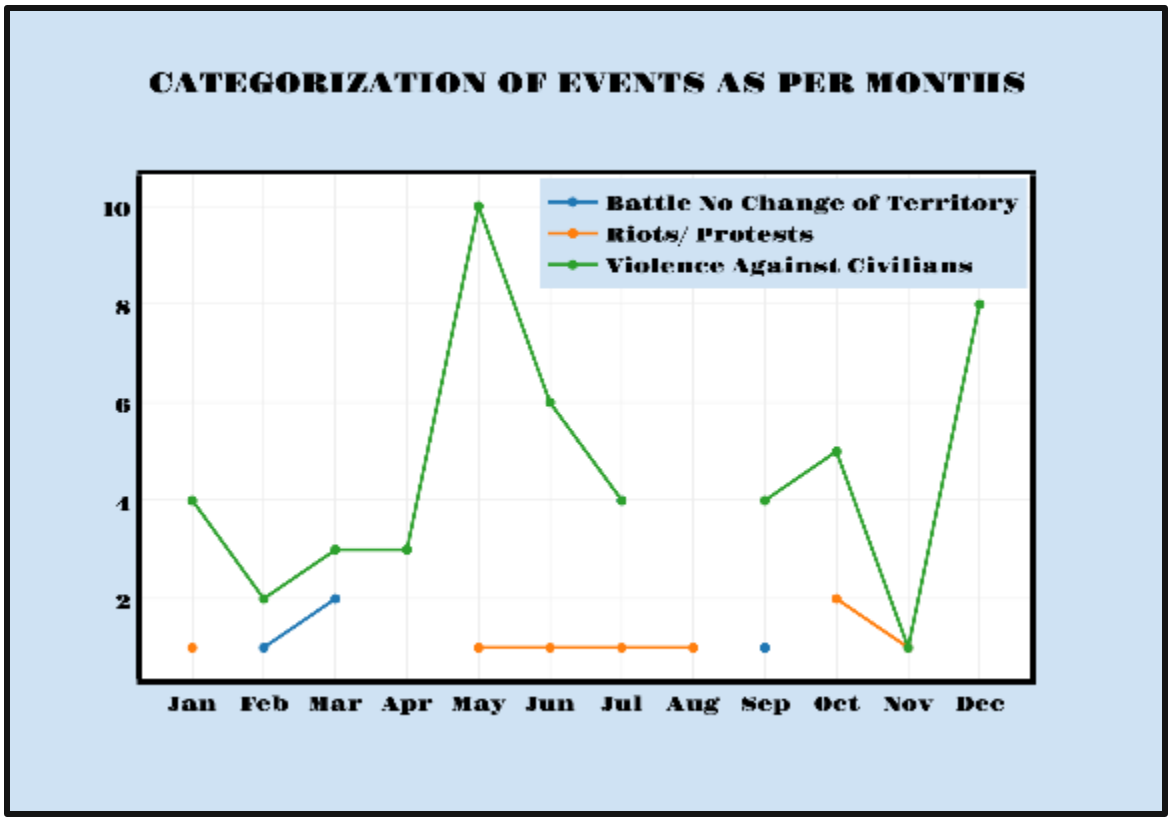
The maximum number of the events in Laos took place in May. The fewest took place in August. In August, the temperature in Laos is hot and humid in comparison to other months. As far as May is concerned, climatic conditions favor violent action since the temperature is relatively moderate.

Table 46. Categorization of Events in Laos per Month

CATEGORIZATION OF EVENTS IN LAOS PER MONTH				
	Battle No Change of Territory	Riots / Protests	Violence Against Civilians	Total
Jan		1	4	5
Feb	1		2	3
Mar	2		3	5
Apr			3	3
May		1	10	11
Jun		1	6	7
Jul	4	1	4	9
Aug		1		1
Sep	1		4	5
Oct		2	5	7
Nov	1	1	1	3
Dec			8	8
Total	9	8	50	67

Most violence against civilians was carried out in May (see Figure 43). In all other months, the number of events that occurred are less than nine.

Figure 43. Categorization of Events per Month



3. Actors Involved in the Conflict

The types of actors involved in Laos conflicts are fewer than in Afghanistan, Cambodia, or other countries involved in irregular warfare. However, the trend of violence against civilians also holds true in this dataset. The second most noticeable trend is the involvement of foreign elements in irregular warfare. In the case of Laos, there was one event involving Muslim militia from Thailand, one event involving military forces from Thailand, and two events involving police forces from Thailand.

Table 47. Relationship Matrix for Laos Actors

RELATIONSHIP MATRIX LAOS												
	Free Democratic People's Government of Laos Militia	Hmong Ethnic Militia (Laos)	Lao Resistance Movement	LPDR: Lao People's Democratic Republic	Military Forces of Laos	Military Forces of Thailand	Muslim Militia (Thailand)	Police Forces of Thailand (2006–2008)	Police Forces of Vietnam (1997–2006)	Protesters	Unidentified Armed Group (Laos)	Grand Total
Civilians	1	6	1	1	11	4	1	1	1		24	51
Free Democratic People's Government of Laos Militia											1	1
Military Forces of Laos	2	3	2									6
Military Forces of Thailand (1997–2001)			1									1
Police Forces of Laos (1998–-2001)										2		2
Police Forces of Thailand (2006–2008)										2		2
Grand Total	2	9	4	1	11	4	1	1	1	8	25	67

4. Geographical Location of Events in Laos

Most events occurred within the geographical limits of Laos. However, a few incidents also took place in Thailand (see Figure 44). This is common to irregular warfare, as a similar pattern was observed in Afghanistan, Pakistan, and other country datasets.

Figure 44. Geographical Location of Events in Laos

F. STATISTICAL ANALYSIS OF DATA FOR MYANMAR

Myanmar, formerly known as Burma, is a country located in Southeast Asia bordering Bangladesh, China, India, Laos, and Thailand. Myanmar gained independence from Great Britain in 1948. Myanmar remained democratic until 1962, when a military coup resulted in the military takeover of the country. The military regime in Myanmar ended in 2011. During the military rule, efforts by the public to restore democracy resulted in political violence. The dataset is focused on events between 1996 and 2009.

1. Statistical Analysis of Data for Myanmar

The dataset for Myanmar includes total of 330 events. Most events (159) are categorized as *battle no change of territory*. The second highest number (104) is categorized as *violence against civilians*.

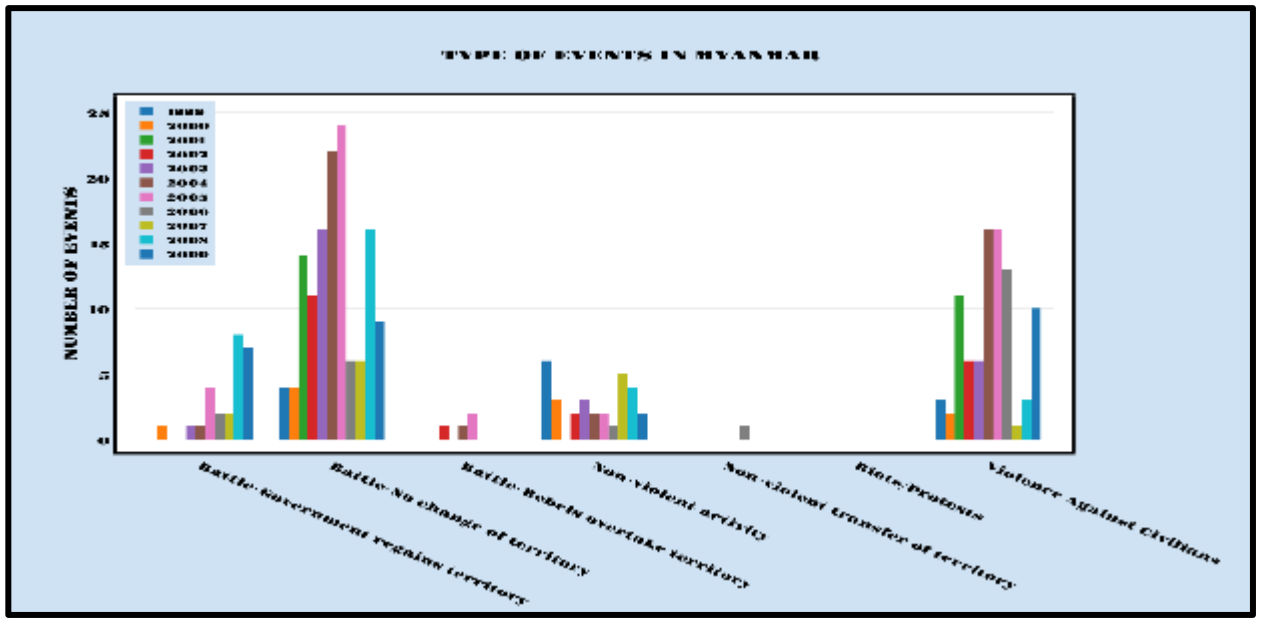
Table 48. Type of Events in Myanmar

TYPE OF EVENTS IN MYANMAR															
Type of Event	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
Battle-Government regains territory		1			1	1	4	2	2	8	7	1	2	1	30
Battle-No change of territory	4	4	14	11	16	22	24	6	6	16	9	9	4	14	159

Battle-Rebels overtake territory				1		1	2						1		5
Non-violent activity by a conflict actor	6	3		2	3	2	2	1	5	4	2				30
Non-violent transfer of territory								1							1
Riots/ Protests												1			1
Violence Against Civilians	3	2	11	6	6	16	16	13	1	3	10	3	2	12	104
Total	13	10	25	20	26	42	48	23	14	31	28	14	9	27	330

Most events took place in 2002, and the second most took place in 2001. The fewest events took place in the year 2008. For the rest of the dataset, the number of incidents are equally distributed over the other years (see Table 48 and Figure 45).

Figure 45. Types of Events in Myanmar from 1996–2009



2. Actors Involved in the Conflict

The relationship matrix of Myanmar involves foreign elements such as military forces from India and Thailand. Besides foreign militaries, there were other foreign actors also involved in the conflict, including the police force of Thailand, the National Socialist Council of Nagaland, and the National Socialist Council of Nagaland, Khaplang. There are 103 recorded events in which civilians were involved. Moreover, civilians in Myanmar emerged as the key players, as in other irregular conflicts (see Table 49).

Table 49. Relationship Matrix for Myanmar Actors

RELATIONSHIP MATRIX MAYANMAR ACTORS																					
	Arakan Liberation Front	Civilians (Myanmar)	Chin National Army	Democratic Karen Buddhist Army	God's Army	Karen National Liberation Army	Karenni National Progressive Party	Military Forces of India	Military Forces of Thailand	Muslim Militia	National Socialist Council Nagaland-Khaplang	National Socialist Council of Nagaland	NUFA: National United Front of Arakan	Police Forces of Myanmar	Police Forces of Thailand	Protesters (Myanmar)	SPDC: State Peace and Development Council	Shan State Army	United Liberation Front from Assam	United Wa State Army	Unknown	Total
ABSDF: All Burma Students Democratic Front		1																			1	2
ALF: Arakan Liberation Front																	1					1
Buddhist Militia										1												1

CNA: Chin National Army		1																				1
DKBA: Democratic Karen Buddhist Army		13				3			5						1							22
God's Army		2																			2	4
Karen National Liberation Army		37		8													22				17	84
Kayin National Union		1																				1
KNPP: Karenni National Progressive Party		1															2				2	5
Military Forces of Bangladesh	1												1									2
Military Forces of India			7																			7
Military Forces of Thailand				1		2												1		1		5

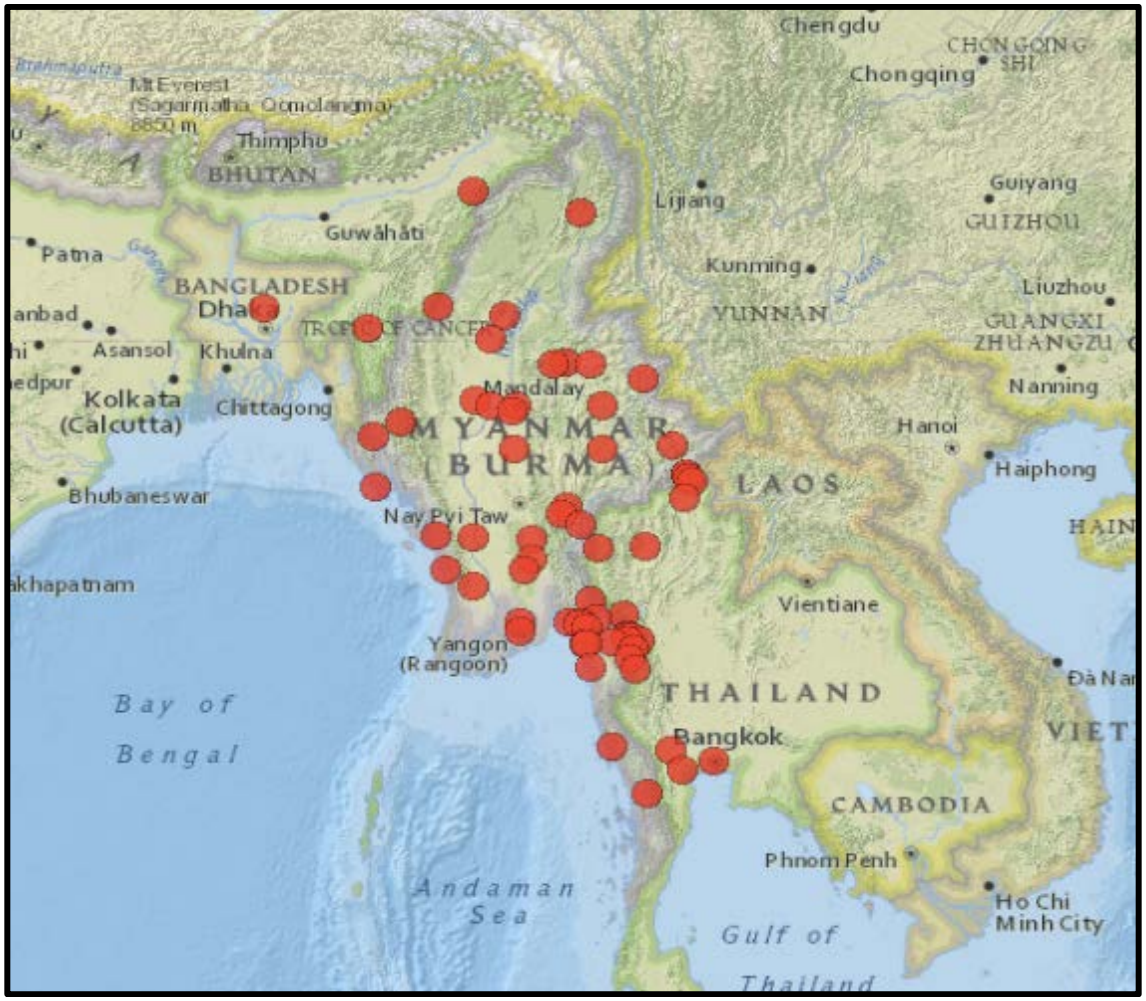
MNDAA: Myanmar National Democratic Alliance Army		2															6					8
NSCN-IM: Nationalist Socialist Council of Nagaland Isaac Chishi Swu and T Muivah																					1	1
NSCN-K: National Socialist Council of Nagaland-Khaplang (India)																					1	1
Police Forces of Myanmar																1						1
Rohingya Solidarity Organisation																	1					1

SPDC: State Peace and Development Council		19			1	3	1	1	3		1	5		1				26	6		1	134
SSA: Shan State Army		3															9				6	18
SSNA: Shan State National Army		2																			1	3
Unidentified Armed Group (Myanmar)		20							1						1		1					23
United Wa State Army		1							2													3
UNLF: United National Liberation Front (India)								1														1
Vigorous Burmese Student Warriors		1																				1
Unknown																						
Grand Total	1	10	7	9	1	4	1	2	1	1	1	5	1	1	2	1	42	27	6	1	32	330

3. Geographical Locations of Event in Myanmar

Most events happened within the geographical boundaries of Myanmar. However, there are a few events that took place in Bangladesh and Thailand (see Figure 46). This pattern is common to other countries where irregular warfare took place.

Figure 46. Geographical Location of Events in Myanmar



G. STATISTICAL ANALYSIS OF DATA FOR NEPAL

Nepal is a landlocked country located in South Asia. Nepal is bordered by China to the north and by India to the south, east, and west. Nepal remained under the Shah dynasty of kings between 1768 and 2008. Civil war broke out between 1996 and 2006. The war resulted in losses to both Maoist and government forces alike, along with heavy losses to civilians. The civil war between Maoist and government forces ended in 2006 with a comprehensive peace accord.

1. Statistical Analysis of Data for Nepal

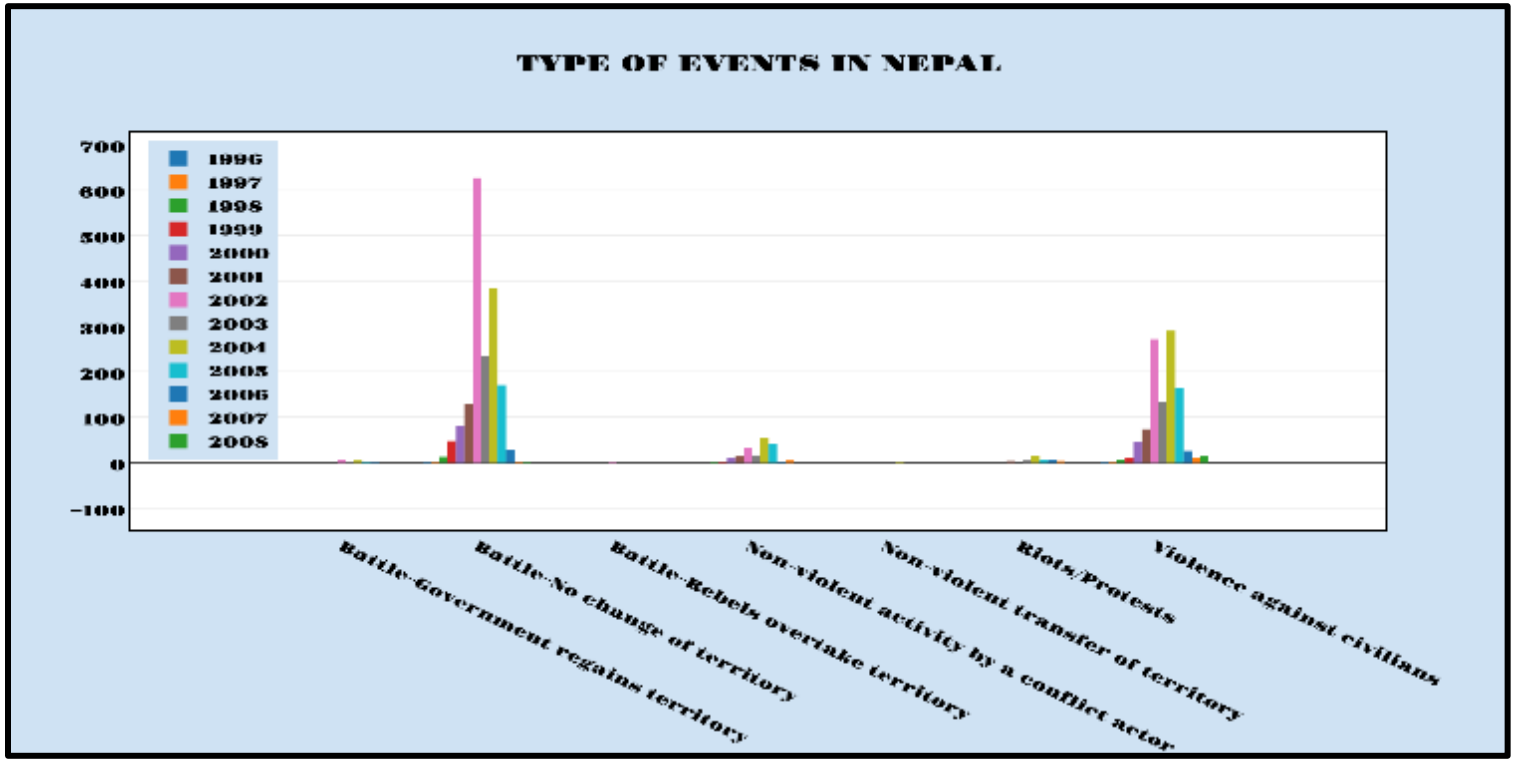
The dataset for Nepal contains seven different types of events. Most events in Nepal are categorized as *battle no change of territory*, for a total of 1724 times. The second highest number is categorized as *violence against civilians* for a total of 1,055 times. The trend of violence against civilians in irregular conflicts is reaffirmed in Nepal's dataset.

Table 50. Type of Events in Nepal

TYPE OF EVENTS IN NEPAL														
Type of Event	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	Total
Battle-Government regains territory							8	2	5	3	1			19
Battle-No change of territory	3	2	13	48	80	129	627	236	383	171	28	2	2	1724
Battle-Rebels overtake territory							2							2
Non-violent activity by a conflict actor			2	2	12	14	33	17	56	41	2	5		184
Non-violent transfer of territory									1					1
Riots/Protests						4		5	15	8	8	4		44
Violence against civilians	1	3	7	10	47	74	272	135	290	164	26	12	14	1055
Total	4	5	22	60	139	221	942	395	750	387	65	23	16	3029

Most events took place in 2002 and 2004. Between1996 and 2004, there was a gradual increase in the number of events. After 2004, there was a gradual decrease in events.

Figure 47. Types of Events in Nepal between 1996 and 2008



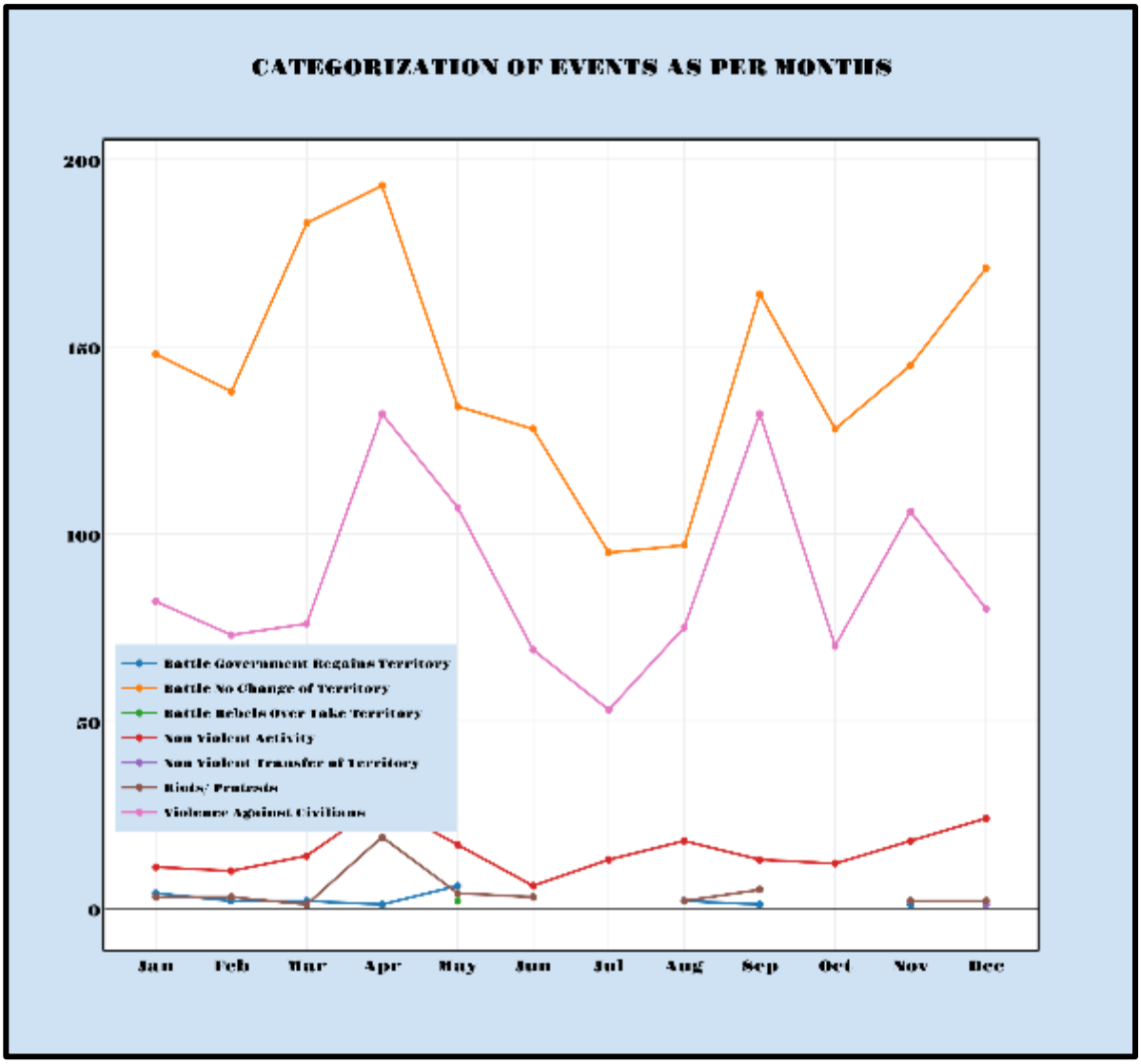
2. Seasonality Effect in Nepal

Most events took place in April. However, from June to August, there is relatively less activity. This is the period in which there is heavy rainfall in Nepal, and the weather is humid and hot (see Table 51 and Figure 48).

Table 51. Categorization of Events in Nepal per Month

CATEGORIZATION OF EVENTS IN NEPAL PER MONTH								
	Battle Government Regains Territory	Battle No Change of Territory	Battle Rebels Over Take Territory	Non Violent Activity	Non Violent Transfer of Territory	Riots/ Protests	Violence Against Civilians	Total
Jan	4	148		11		3	82	248
Feb	2	138		10		3	73	226
Mar	2	183		14		1	76	276
Apr	1	193		28		19	132	373
May	6	134	2	17		4	107	270
Jun		128		6		3	69	206
Jul		95		13			53	161
Aug	2	97		18		2	75	194
Sep	1	164		13		5	132	315
Oct		128		12			70	210
Nov	1	145		18		2	106	272
Dec		171		24	1	2	80	278
Total	19	1724	2	184	1	44	1055	3029

Figure 48. Categorization of Events per Month



3. Actors Involved in Conflict

The numbers of actors involved in Nepal's conflict were fewer than those of other conflicts. Another important factor, which differentiates Nepal from the rest of the countries, is that there were no foreign elements involved in the conflict. The communist force of Nepal emerged as the biggest player in the highest number of conflicts. After the communist party of Nepal, civilians emerged as the biggest players. The most conflicts involved the Maoist force against the military force of Nepal.

Table 52. Relationship Matrix Nepal Actors

RELATIONSHIP MATRIX NEPAL ACTORS										
	Civilians (Nepal)	Communist Party of Nepal-Maoist	Madeshi People's Rights Forum	Madhesi Tigers	Military Forces of Nepal	Police Forces of Nepal	Protesters (Nepal)	Unidentified Armed Group (Nepal)	Unknown	Grand Total
Civilians (Nepal)						1				1

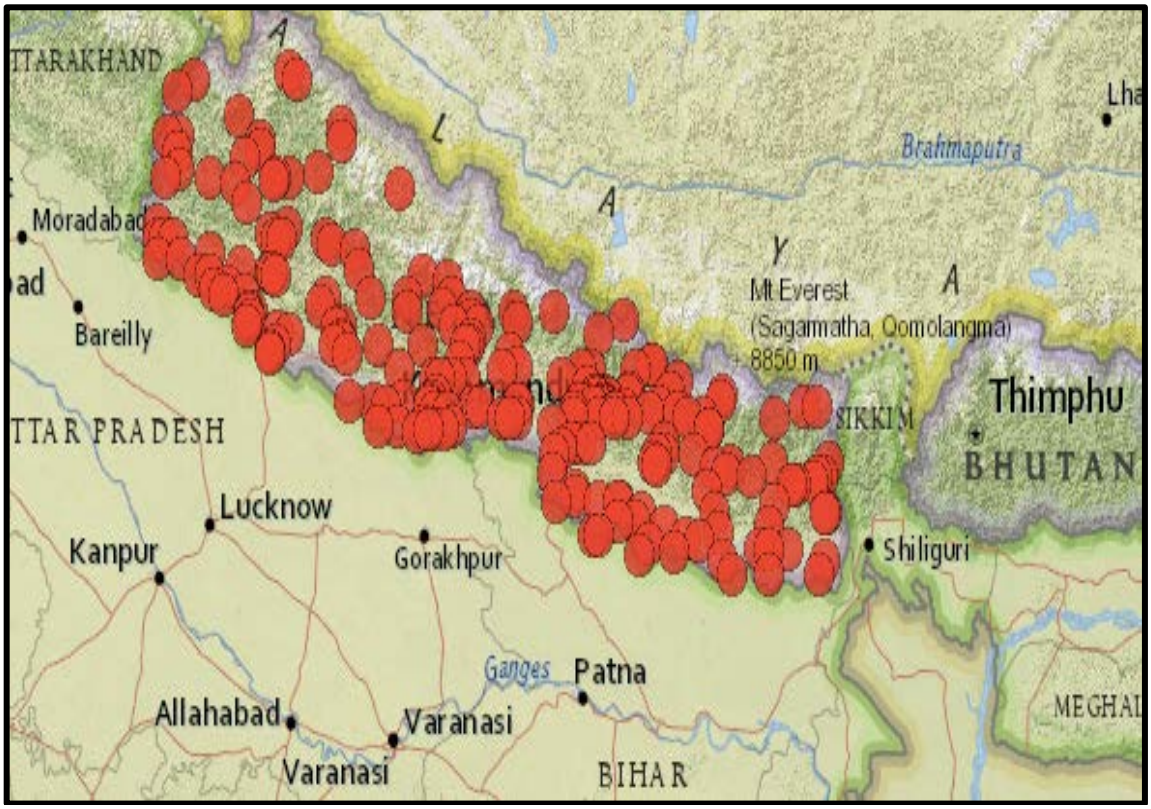
Communist Party of Nepal-Maoist	964		1	1	362	323		1	189	1841
Madhesi Tigers									1	1
Military Forces of Nepal	35	955							4	994
Newa National Liberation Front									1	1
PLF: People's Liberation Front	1					1				2
Police Forces of Nepal	26	75			1		13			115
Protesters (Nepal)		2			3	17			5	27
Rioters (Nepal)						2			1	3
Royal Nepalese Army		1								1
Terai Army	3									3

Unidentified Armed Group (Nepal)	24	6			1	1				32
Unknown	1	4				1			2	8
Total	1054	1043	1	1	367	346	13	1	203	3029

4. Geographical Locations of Event in Nepal

In the case of Nepal, most of the events were widely spread across the country. However, another trend, which was identified in Afghanistan, Cambodia, and other countries, was that the frequency of events is more in close vicinity to international borders (see Figure 49). Most incidents happened in the area bordering India. There were less events in the areas bordering China.

Figure 49. Geographical Location of Events in Nepal



H. STATISTICAL ANALYSIS OF DATA FOR PAKISTAN

The global war on terrorism has significantly impacted Pakistan. War drastically impacted internal security and left both short- and long-term imprints on the socio-economic dynamics of Pakistani society. The Afghan war's proximity to Pakistan resulted in the emergence of extremist and radical Jihadi organizations within Pakistan. This spread of radicalization engulfed society, resulting in chaos and widespread terrorism. Terrorism coupled with sectarian conflicts resulted in heavy losses to life and material. Pakistan's dataset is not a complete picture of the conflict; it only contains data from 2005 to 2009. These five years of data include 4,759 events, which are higher than any other country in the dataset.

1. Statistical Analysis of Data for Pakistan

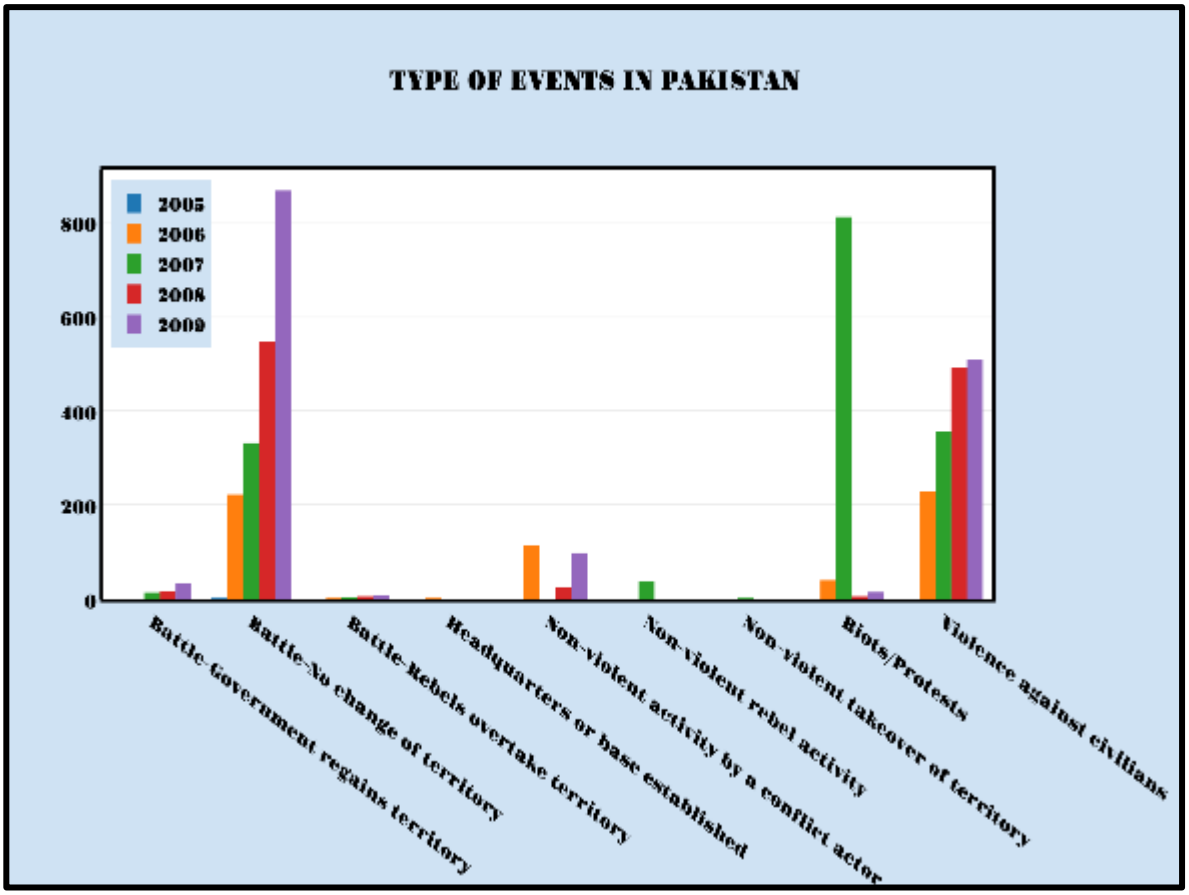
Pakistan's dataset contains all nine types of events. The maximum number of events happened in 2007. In 2009, there were 1,523 events. According to the dataset, only one event happened in 2005; this may not be correct. There is a possibility that data for Pakistan might not have been recorded or reflect only a small portion of the year. Most events (1,966) fall within the *battle no change of territory* category. The second most common event is *violence against civilians*, which happened 1,580 times (see Table 53).

Table 53. Type of Events in Pakistan

TYPE OF EVENTS IN PAKISTAN						
Type of Event	2005	2006	2007	2008	2009	Total
Battle-Government regains territory			13	15	32	60
Battle-No change of territory	1	221	331	547	866	1966
Battle-Rebels overtake territory		1	3	5	6	15
Headquarters or base established		1				1
Non-violent activity by a conflict actor		112		23	97	232
Non-violent rebel activity			36			36
Non-violent takeover of territory			1			1
Riots/Protests		39	810	5	14	868
Violence against civilians		228	354	490	508	1580
Total	1	602	1548	1085	1523	4759

Because data from Pakistan spans only five years, we cannot determine any significant trend. The only observation that can be made is that incidents have generally increased from 2005 to 2009.

Figure 50. Types of Events in Pakistan from 2005–2009



2. Seasonality Effect in Pakistan

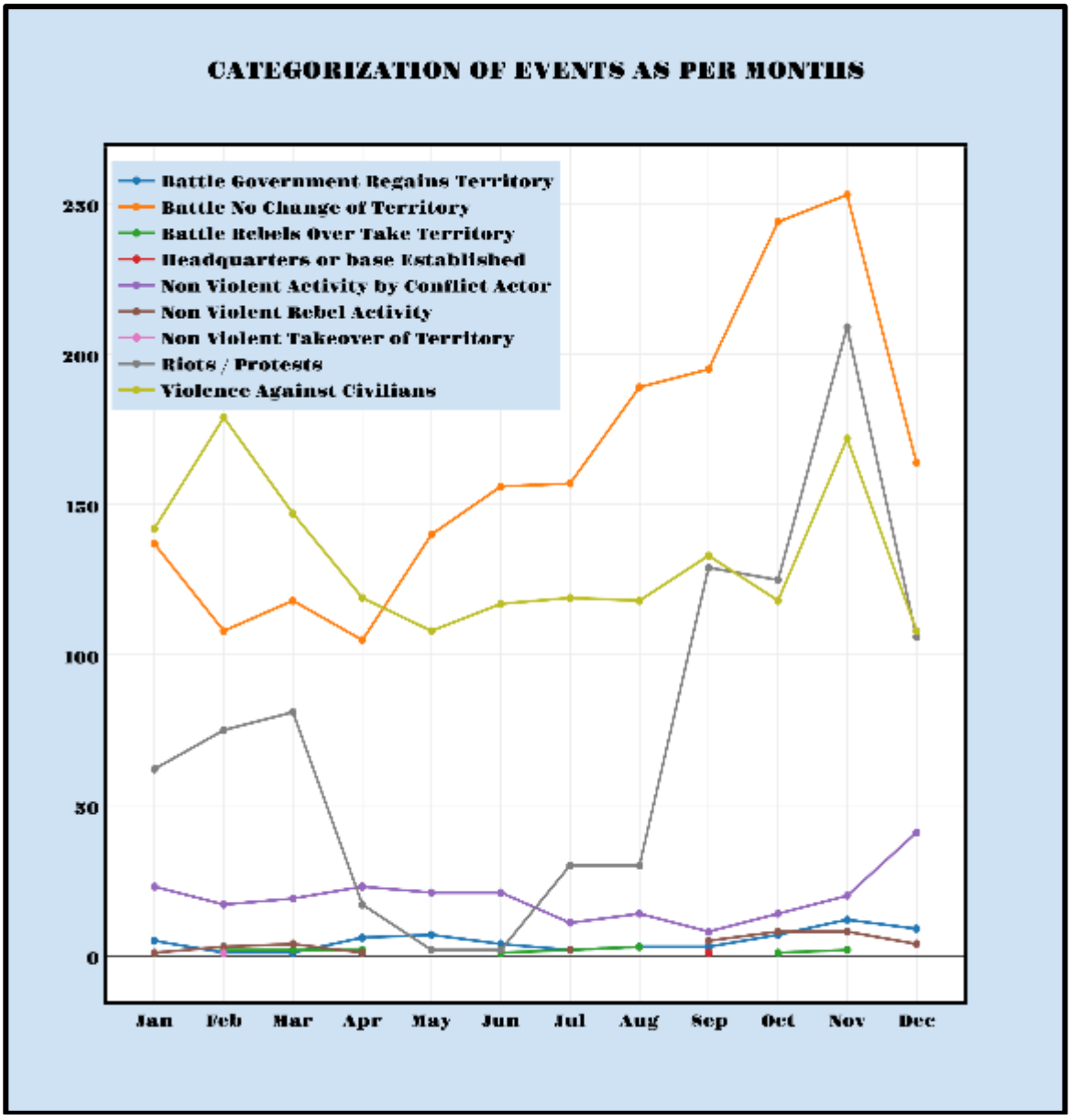
The seasonality effect cannot be exactly determined because events were widely spread over the entire country. Moreover, Pakistan has four distinct seasons and different weather conditions in each region. To more precisely model the seasonality effect in Pakistan, regions and the number of events in each region are analyzed separately. For all of Pakistan, the most events took place between October and November. In these months, the event *battle no change of territory* and *riots/protests* happened with the highest frequency, relatively. This trend can be attributed to weather conditions because weather in these two months in all regions is fairly pleasant (see Table 54).

Table 54. Categorization of Events in Pakistan per Month

CATEGORIZATION OF EVENTS IN PAKISTAN PER MONTH										
	Battle Government Regains Territory	Battle No Change of Territory	Battle Rebels Over Take Territory	Headquarters or base established	Non Violent Activity by Conflict Actor	Non Violent Rebel Activity	Non Violent Takeover of Territory	Riots/ Protests	Violence Against Civilians	Total
Jan	5	137			23	1		62	142	370
Feb	1	108	2		17	3	1	75	179	386
Mar	1	118	2		19	4		81	147	372
Apr	6	105	2		23	1		17	119	273
May	7	140			21			2	108	278
Jun	4	156	1		21			2	117	301
Jul	2	157	2		11	2		30	119	323
Aug	3	189	3		14			30	118	357
Sep	3	195		1	8	5		129	133	474
Oct	7	244	1		14	8		125	118	517
Nov	12	253	2		20	8		209	172	676
Dec	9	164			41	4		106	108	432
Total	60	1966	15	1	232	36	1	868	1580	4759

From October to November, there is an increase trend of events (see Figure 51). From April to August, there are relatively fewer events.

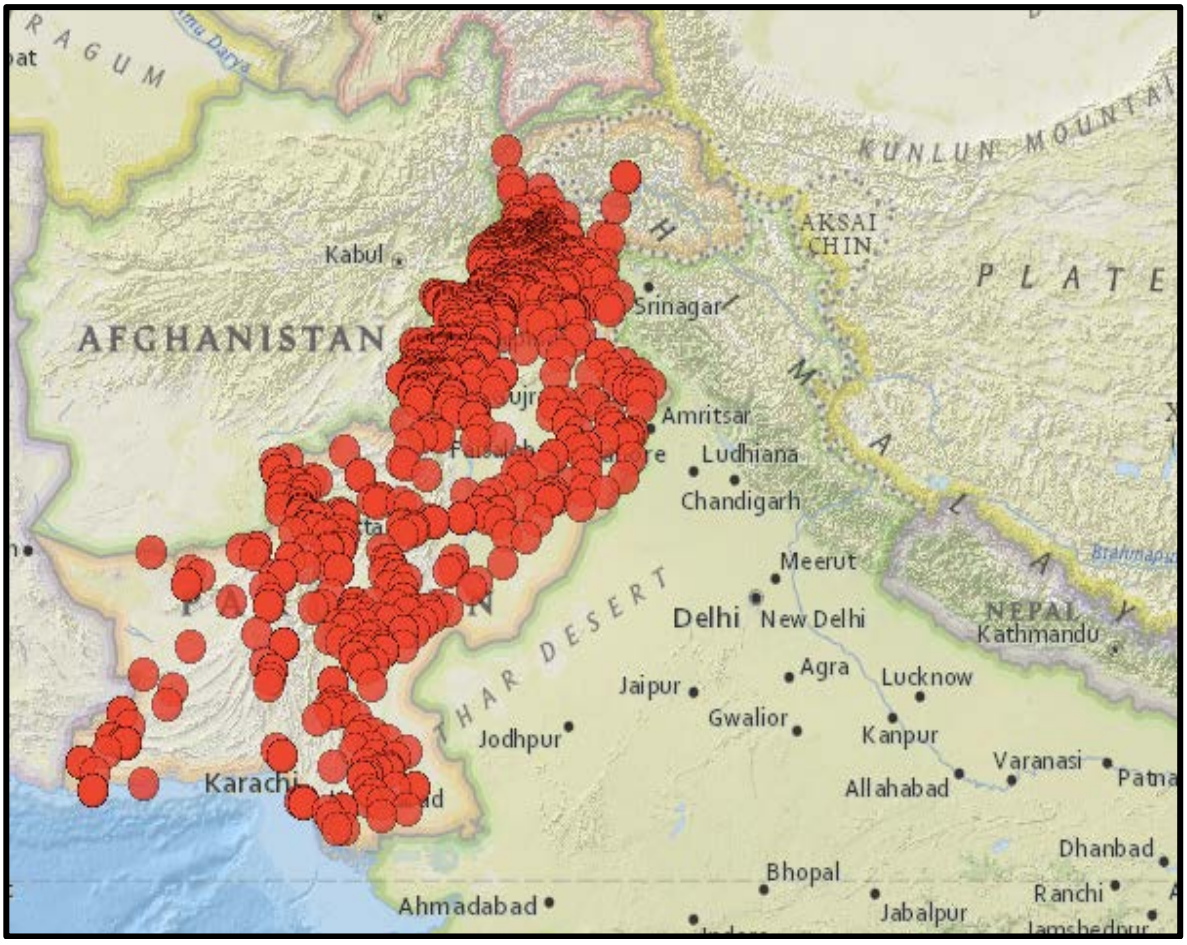
Figure 51. Categorization of Events per Month



3. Geographical Location of Events in Pakistan

The geographical map for Pakistan shows widespread events across all regions. The most populated events occurred in the northwestern portion of Pakistan known, as Federal Administrated Tribal Areas (FATA). These areas have been in the limelight since the commencement of the global war on terrorism.

Figure 52. Geographical Location of Events in Pakistan



I. STATISTICAL ANALYSIS OF DATA FOR AFRICA

Africa is the second largest and second most populous continent. Africa is bordered to the north by the Mediterranean Sea; to the northeast by the Suez Canal, Red Sea, and Sinai Peninsula; to the southeast by the Indian Ocean; and to the west by the Atlantic Ocean. Africa comprises 55 recognized states, of which 54 are members of the African Union. Africa remained under European colonialism into the twentieth century, which resulted in underdevelopment of the continent compared to the rest of the world. Underdevelopment combined with an abundance of natural resources on the continent has resulted in large-scale political violence and conflict. The dataset contains information for 49 countries of the continent, but it does not have information regarding the islands of Cape Verde, Comoros, Sao Tome and Principe, and Seychelles. A holistic

analysis of Africa is carried out in this research, and countries are not segregated for analysis. Therefore, the seasonality effect is not modeled for Africa.

1. Number of Events in Africa

The ACLED dataset for Africa includes 25 variables and 99,549 events. Its data span events between 1997 and 2014. Unlike the datasets for the aforementioned countries in this chapter, this dataset includes the number of fatalities. Nevertheless, this dataset is organized after the same pattern as the other ACLED dataset. The numbers of events, by country, are given in Table 55. The top five countries in terms of event frequency are Somalia; the Democratic Republic of Congo; Nigeria; Sudan; and Egypt; with 15,150; 8,876; 6,781; 6,505; and 5,739 incidents of violence; respectively. These totals are much higher than those recorded for other countries. The five countries ranking lowest in terms of violence are Benin, Gambia, Djibouti, Botswana, and Equatorial Guinea, with 100, 95, 91, 43, and 38 incidents, respectively. The mean for the number of events in Africa's dataset is 1,990.9; the median is 670.5; the standard deviation is 2,847.7; and the standard error is 402.7.

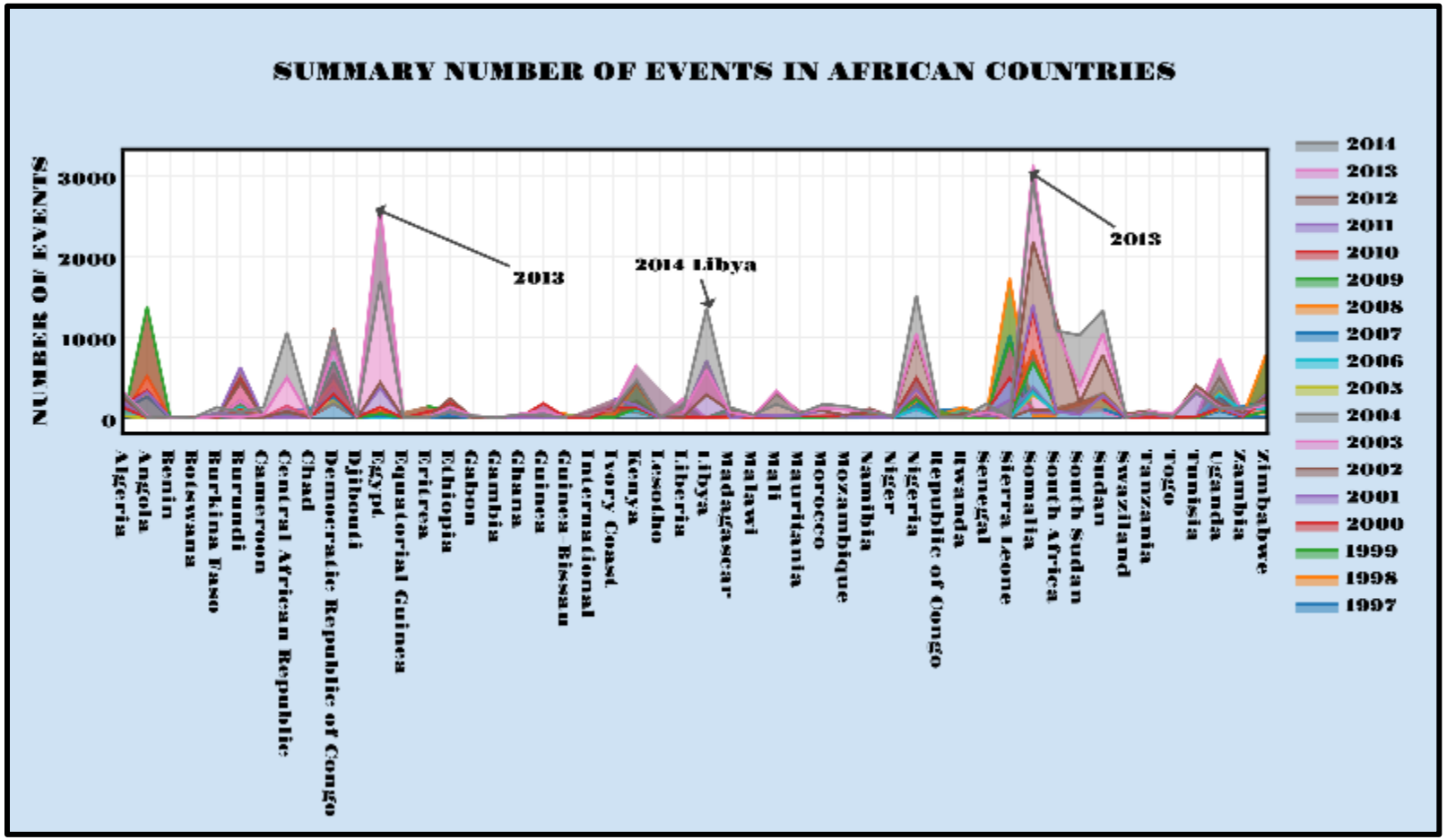
Table 55. Summary Number of Events in African Countries

SUMMARY NUMBER OF EVENTS IN AFRICAN COUNTRIES																			
Countries	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Total
Algeria	115	49	35	170	124	41	49	105	27	202	317	237	283	124	194	234	314	289	2909
Angola	264	523	1379	345	323	27	7	7	3	6	15	10	13	9	20	16	25	9	3001
Benin	2	4	2	1	3	1	2	5	2	1	3	1	2	10	12	18	13	18	100
Botswana	1	2	1	1			2	10	3	2	2	2	3	3	3	2	1	5	43
Burkina Faso	4	5	33	22	6	7	1	4	6	8	1	12	1	6	63	17	25	139	360
Burundi	217	116	165	505	631	440	216	78	74	133	29	104	76	102	30	20	13	79	3028
Cameroon	14	6	2	8	4	2	5	15	19	6	20	23	19	15	13	13	43	135	362
Central African Republic	41	20	14	12	41	59	67	31	19	48	100	85	95	156	14	84	503	1061	2450
Chad	15	12	12	22	11	15	14	36	43	119	103	76	23	11	6	7	6	27	558
Democratic Republic of Congo	293	305	558	178	349	439	274	176	308	332	267	446	705	304	891	1116	837	1098	8876
Djibouti	1	2	7	1	2	1	1	2	3	3	2	2	6	10	6	2	12	28	91
Egypt	43	19	10	49	25	51	15	11	62	13	36	92	53	134	370	444	2617	1695	5739
Equatorial Guinea	3	2		2	1	1	1	3	2	1	2	2	8	1	1	3	1	4	38
Eritrea	2	62	157	79	3	2	8	9	5	10	5	21	7	3	5	7	7	4	396
Ethiopia	32	68	76	137	59	240	116	124	129	91	49	112	105	190	90	162	139	71	1990
Gabon	1	2	3	2	4	1	1	3	1	4		2	8	3	4	7	32	29	107
Gambia	2	1	4	8	14	13	11	5	3	3	2	4	2	4	4	3	5	7	95
Ghana	5	4	2	10	9	7	17	7	3	3	5	11	17	14	12	48	53	50	277
Guinea	3	34	45	184	71	14	10	12	15	8	47	15	41	25	10	38	94	39	705
Guinea-Bissau	5	64	18	9	3	5	5	7	14	21	2	3	9	4	9	18	8	5	209
International							1		3		1		1	1					7
Ivory Coast	17	18	81	177	38	138	181	105	50	32	6	25	5	133	240	85	115	118	1564
Kenya	117	177	211	200	166	198	84	72	153	103	434	418	127	131	158	436	668	467	4320
Lesotho	4	35	4	3	2	2	4	1	1	3	2	2	2	3	4	20		12	104

Liberia	30	30	45	89	179	149	242	26	30	22	9	17	9	13	32	21	70	122	1135
Libya	2	10	3	5	4	5	3	7	7	7	1	7	6	7	715	292	594	1367	3042
Madagascar	5	4	1	4	6	116	18	22	11	10	17	1	103	17	3	91	73	134	636
Malawi	1	5	2	16	23	11	15	5	6	6	6	5	4	4	35	15	24	40	223
Mali	15	4	10	7	7	2	4	4	2	10	11	21	10	20	29	300	339	171	966
Mauritania	5	1	3	4	1	3	13	3	9	3	4	15	10	22	38	82	68	61	345
Morocco	14	14	20	24	15	16	21	8	19	16	21	16	11	22	99	91	123	173	723
Mozambique	12	1	4	11	17	3	28	32	20	7	11	21	49	29	18	39	110	149	561
Namibia	8	4	55	76	20	7	13	10	9	12	19	20	10	12	16	118	88	76	573
Niger	34	44	10	13	9	42	6	18	9	8	31	28	16	12	10	12	26	27	355
Nigeria	153	159	209	171	105	153	202	271	202	114	188	207	232	504	349	991	1043	1528	6781
Republic of Congo	95	42	76	6	5	45	3	7	10	1	3		9	5	1	7	4	13	332
Rwanda	125	138	13	19	27	5	4	2	5	12	20	12	17	54	41	56	21	12	583
Senegal	52	34	33	86	61	40	19	11	11	29	24	20	22	24	36	63	74	176	815
Sierra Leone	1026	1735	945	499	225	5	20	14	5	1	3	14	21	11	9	10	10	21	4574
Somalia	25	27	86	103	66	110	387	335	288	320	847	843	694	1325	1406	2192	3140	2956	15150
South Africa	48	37	58	53	61	96	55	45	101	57	104	91	143	65	86	1243	1111	1086	4540
South Sudan															44	192	382	1025	1643
Sudan	211	213	187	205	205	328	70	217	315	154	115	303	289	234	285	788	1054	1332	6505
Swaziland	7	7	2	8	2	5	8	2	5	5	4	12	1	7	11	55	22	12	175
Tanzania	10	5	8	17	34	24	8	37	68	16	6	26	9	5	53	94	77	67	564
Togo	9	13	3	4	4	3	4	3	25	1	1	3	4	18	13	20	58	14	200
Tunisia	3	2	3	4	5	8	5	6	10	1	2	1	1	19	310	405	341	327	1453
Uganda	92	345	261	275	140	500	739	387	305	296	104	111	131	124	138	181	132	120	4381
Zambia	13	18	15	22	19	4	50	31	61	95	143	24	51	18	37	63	108	117	889
Zimbabwe	22	91	83	341	318	784	353	324	273	122	207	793	225	225	298	262	160	194	5075
Total	3218	4513	4954	4187	3447	4168	3382	2655	2754	2477	3351	4316	3688	4192	6271	10483	14783	16709	99548

Figure 53 depicts trends by country over the time span. In terms of event frequency, 2014, then 2013, are the most violent years. The graph shows that Somalia, the Democratic Republic of Congo, Nigeria, Sudan, and Egypt emerge as the countries experiencing the most violence over the past two decades.

Figure 53. Number of Events in African Countries



2. Type of Events in Africa

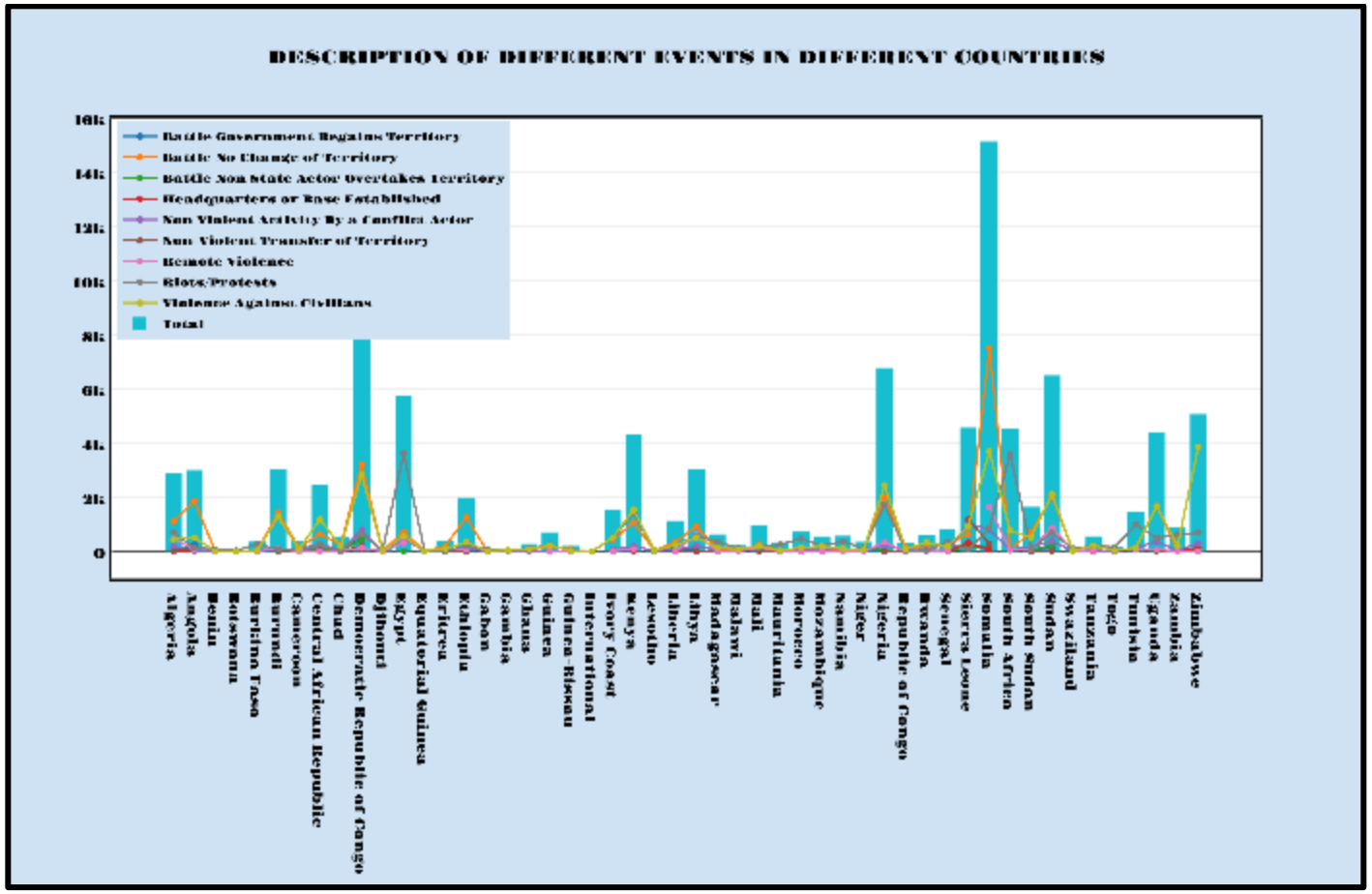
Table 56 indicates all nine types of events in African countries. The most common event that occurred was *battle no change of territory*, which happened 30,131 times. *Violence against civilians* ranks second in frequency, occurring in 29,001 events. The third most common event was the riots/protest category, with 22,267 events.

Table 56. Summary of Type of Events in African Countries

SUMMARY OF TYPE OF EVENTS IN AFRICAN COUNTRIES																			
Events	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Total
Battle-Government regains territory	156	200	208	84	91	75	64	19	41	68	37	65	77	36	82	100	110	149	1662
Battle-No change of territory	894	1263	2324	1669	1295	1381	1161	894	746	857	1274	1581	1129	1420	1714	2330	3804	4395	30131
Battle-Non-state actor overtakes territory	156	172	156	66	96	203	121	19	10	49	21	41	51	12	132	112	100	144	1661
Headquarters or base established	101	133	63	88	27	42	26	2	1	8	5	13	13	49	40	36	46	30	723
Non-violent activity by a conflict actor	352	444	419	211	212	126	92	99	107	103	88	181	261	279	551	1208	929	1014	6676
Non-violent transfer of territory	214	636	308	202	95	30		4	8	19	17	41	40	20	60	93	70	152	2009
Remote violence	102	109	163	178	108	118	149	131	22	85	153	131	196	301	575	787	863	1247	5418
Riots/Protests	339	392	405	439	324	437	371	379	714	432	618	537	690	733	1478	3281	5359	5339	22267
Violence against civilians	904	1164	908	1250	1199	1756	1398	1108	1105	856	1138	1726	1231	1342	1639	2536	3502	4239	29001
Total	3218	4513	4954	4187	3447	4168	3382	2655	2754	2477	3351	4316	3688	4192	6271	10483	14783	16709	99548

Figure 54 shows trends of different events by country. Somalia, the Democratic Republic of Congo, Nigeria, Sudan, and Egypt have the most events, as indicated by the blue bars and lines for event frequency

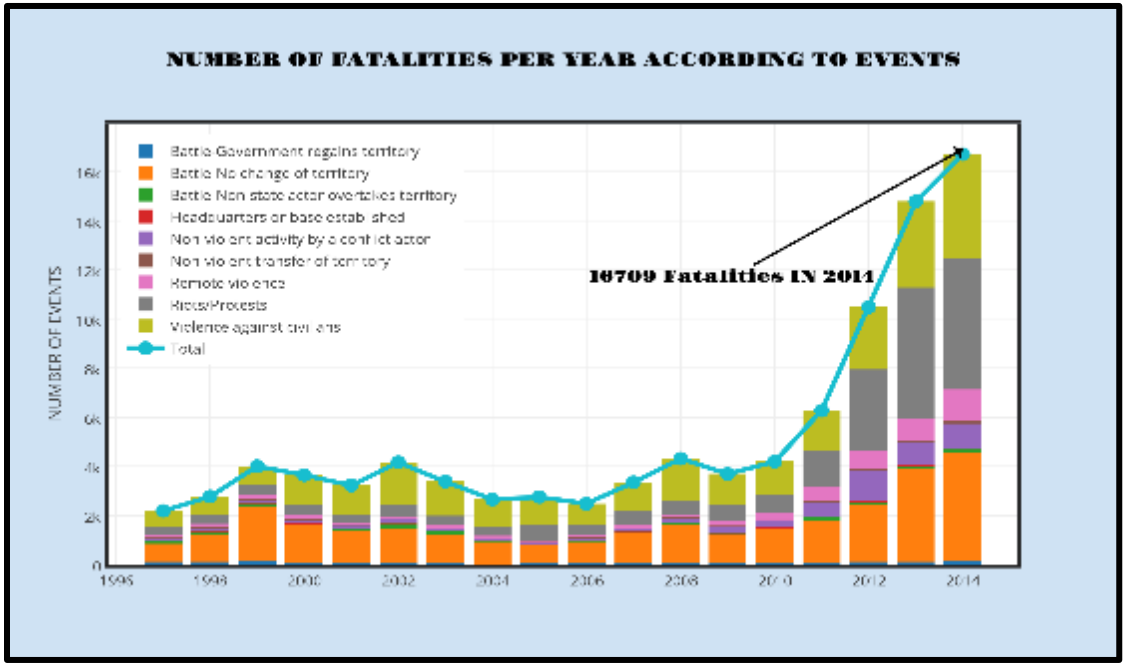
Figure 54. Description of Different Events in African Countries



3. Number of Fatalities

Figure 55 indicates the fatalities according to different events between 1996 and 2014. The orange color in the bar reflects the fatalities due to battle no change of territory, gray reflects battle due to riots, and light green is violence against the civilians. The most fatalities, which are 30,131, occurred because of *battle no change of territory*; 29,001 fatalities were because of *violence against civilians;* and 22,267 fatalities occurred because of *riots/protests*. There is an increased trend of violence in Africa.

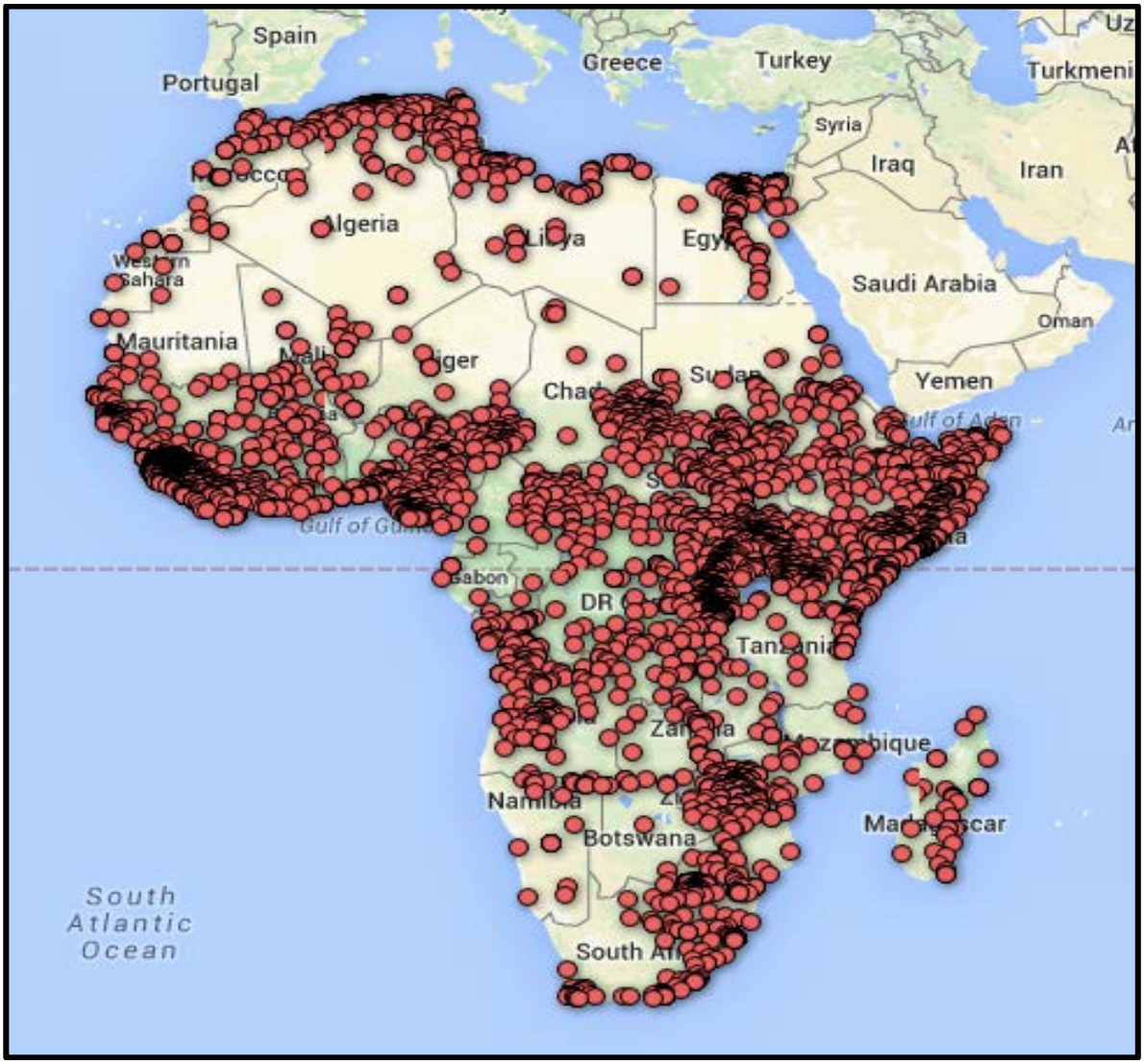
Figure 55. Number of Fatalities per Year According to Events



4. Geographical Location of Events in Africa

Event density in Africa is given in Figure 56, reflecting 99,548 total events.

Figure 56. Geographical Location of Events in Africa



V. STATISTICAL ANALYIS OF CORRELATES OF WAR AND UPPSALA DATASETS

A. STATISTICAL ANALYSIS OF CORRELATES OF WAR PROJECT DATASET

Professor J. David Singer from the University of Michigan founded the Correlates of War (COW) project in 1963, with a goal of carrying out a scientific analysis of warfare. Singer later partnered with Melvin Small to collect precise data on the wars related to the post-Napoleonic period.[115] In 1972, Singer and Small published *The Wages of War* and defined terminologies on warfare.[116] This initial work became the theoretical basis for numerous studies on warfare. The aim of the COW project was to identify variables that systematically explained temporal and spatial variations in warfare. Initially, efforts were made to measure variables that ostensibly resulted in war, such as national capacity, treaties and alliances, and location in the post-Napoleonic era.

The project has expanded over the years, beyond data collection, to include various empirical studies about warfare. According to Singer and Small's definition, war involves violence, and the dominant characteristic of war is brutality.[117] Based on this definition, they measure war against two criteria: *battle related casualties* and *participants on two sides*. Singer and Small declared 1,000 battle-related casualties as the threshold for differentiating war from other conflicts.[118] Initially, Singer and Small defined war topology by type of participants and focused on inter-state system or states. According to them, wars can be divided into international wars and civil wars.[119] They further subdivided international war into two categories: *inter-state* and *extra-systemic*. C*ivil war* was defined as military action internal to the state system with active participation of national government and resistance from both sides.

---

115 David Joel Singer and Small Melvin. *The Wages of War, 1816–1965: A Statistical Handbook.* (Hoboken, NJ: John Wiley & Sons, 1972).

116 Ibid.

117 Meredith Reid Sarkees, "The COW Typology of War: Defining and Categorizing Wars (Version 4 of the Data)," *Note with Version 4 of the Correlates of War Data* (2010a).

118 Ibid.

119 Ibid.

In 1994, a new topology was introduced by the COW project to accommodate additional types of armed conflict that fell outside the existing categories. The primary addition to the COW project was incorporation of non-state actors.[120] The details of the two topologies are given in Table 57. In this chapter, datasets from the COW project related to intra-state wars are examined, since we are only considering irregular warfare in this chapter.

Table 57. COW Project's Typologies of War[121]

TYPOLOGIES OF WAR									
TRADITIONAL TYPOLOGY					EXPANDED TYPOLOGY				
I.	International wars				I.	Inter-state wars			
	A.	Inter-state wars			II.	Extra-state wars			
	B.	Extra-systemic wars				A.	Colonial-conflict with colony		
		1.	Colonial			B.	Imperial state vs. Non-state		
		2.	Imperial		III.	Intra-state wars			
II.	Civil wars					A.	Civil wars		
							1.	For central control	
							2.	Over local i\ssues	
						B.	Regional internal		
						C.	Inter-communal		
					IV.	Non-state wars			
						A.	In non-state territory		
						B.	Across state borders		

B. STATISTICAL ANALYSIS OF COW INTRA-STATE WAR

The COW project breaks intra-state war into three sub categories: civil, regional internal, and inter-communal wars. Civil war is between a defined government of state and a non-state entity. Civil war is further subdivided into two types: for control of the central government and for local issues.[122] Regional war is between a government of a regional subunit and a non-state entity. Inter-communal war involves combat between

120 Ibid.
121 Ibid.
122 Ibid.

two or more non-state entities within the state. A central government has de facto control over territory regardless of its legality. The dataset contains wars between 1818 and 2007. The first war in the dataset is the First Caucasus, fought in 1818. The last battle in the dataset is the second Yemen cleric fight between Yemen and Zaidi Muslims in 2007.

1. Types of Wars and Regions

Most intra-state wars fall within the category of *civil war for central control*, with a total of 250. There are 161 wars within *civil war over local issues*, 11 wars within *regional internal war*, and 20 wars within *internal communal war*. The intra-state dataset divides wars into six regions: the Western Hemisphere, Europe, Africa, the Middle East, Asia, and Oceania. The highest numbers of wars were fought in Africa. These results are close to what we observed in the ACLED dataset in which most violence was also carried out in Africa. After Africa, Asia ranks second in terms of intra-state war. A detailed summary of the different types of war and regions in which they were fought is given in Table 58.

Table 58. Types of Wars in COW Intra-State War Dataset

TYPOLOGIES OF WAR							
Type of War	W. Hemisphere	Europe	Africa	Middle East	Asia	Oceania	Total
Civil War for Central Control	62	40	73	31	44	-	250
Civil War over Local Issues	17	48	24	27	45	-	161
Regional Internal War	2	1		6	2	-	11
Inter-communal		2	7	8	3	-	20
Total	81	91	104	72	94	-	442

2. Outcome of War

The COW's intra-state war dataset is partitioned into seven categories. These categories include winning by side A (government), winning by side B (insurgents/non-state groups), compromise between two sides, transformation of war into another type, continuity of conflict, stalemate, and continuity of war below war level. Side A is government and side B is mostly insurgents or non-state groups. According to the dataset, 221 wars were won by side A, 105 by side B, 45 wars resulted in compromise, 2 conflicts are ongoing, 24 resulted in stalemate, and 17 continue below war level. The two conflicts that continue today include the Eighth Colombia War between Colombia and FARC, drug lords, and the Second Sri Lanka Tamil war between Sri Lanka and the Liberation Tigers of Tamil Eelam (LTTE). However, this is not accurate, because LTTE was defeated by Sri Lankan armed forces in 2009.[123]

Table 59. Types of Wars and Outcomes

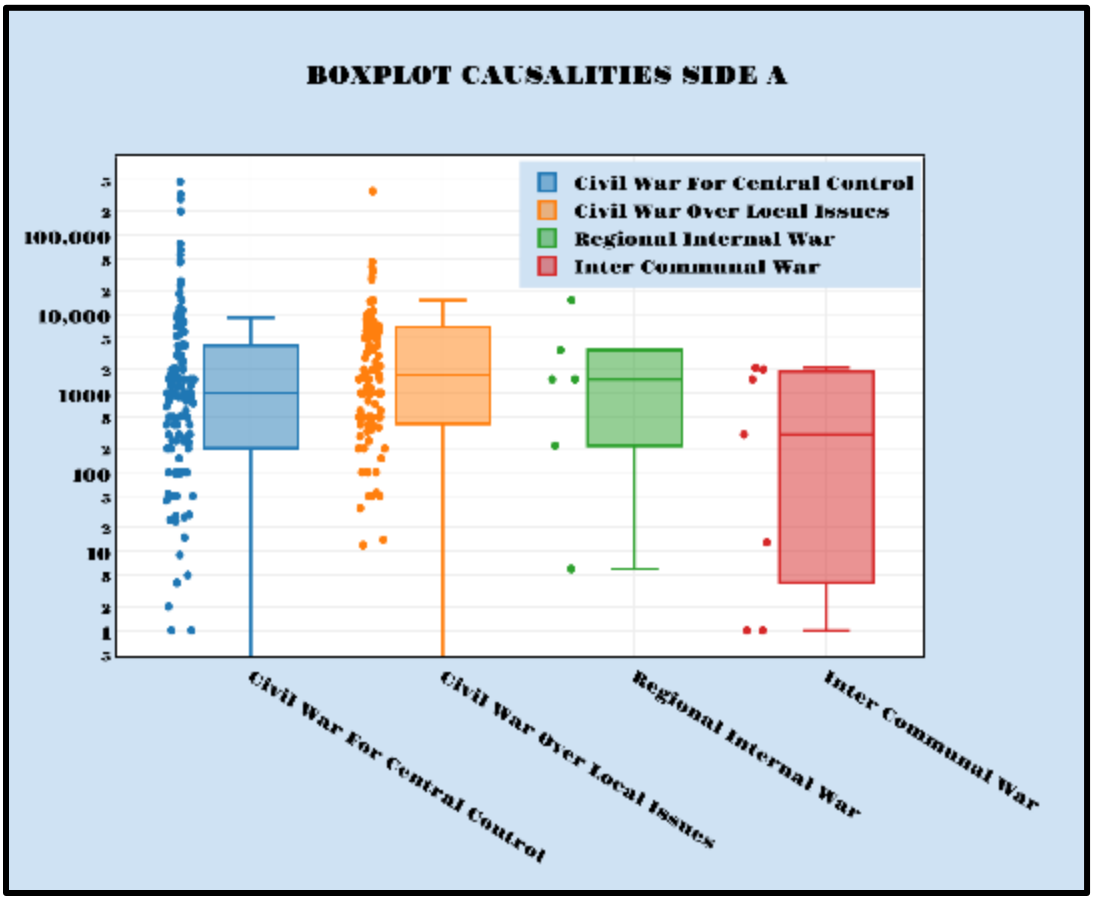
TYPOLOGIES OF WAR					
Type of War	Civil War for Central Control	Civil War over Local Issues	Regional Internal War	Inter-communal	Total
Side A (Government) Wins	116	100	2	3	221
Side B (Insurgents) Wins	76	19	6	4	105
Compromise	34	9		2	45
Transformed	10	15	3		28
Ongoing 12/31/ 2007		2			2
Stalemate	7	9		8	24
Conflict Continues Below War level	7	7		3	17
Total	250	161	11	20	442

[123] Neil DeVotta, "The Liberation Tigers of Tamil Eelam and The Lost Quest for Separatism in Sri Lanka," Asian Survey (2009): 1021–1051.

3. Number of Casualties

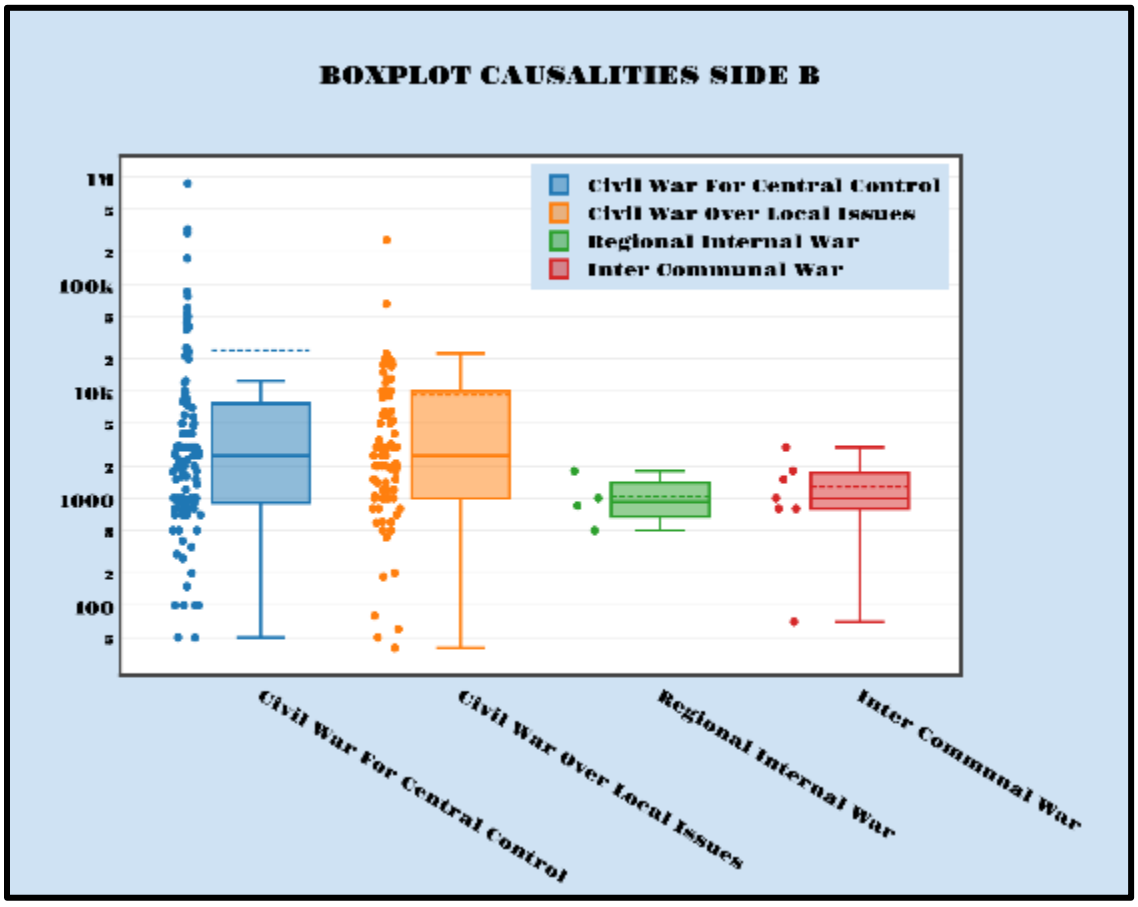
There are 171 wars in which the number of casualties for side A is unknown, and 42 wars in which the number of casualties for side B is unknown. The greatest number of casualties suffered by side A was 475,000 in the Russian civil war from 1917 to 1921. This war falls under the category of civil war for central control. There were two wars that had zero casualties for side A: (1) the First Liberia war, fought during 1990 between Liberia and the National Patriotic Front of Liberia (NPFL), and (2), the Bosnian-Serb Rebellion during 1995 between the United States and Bosnian Serbs. In this war, other allied countries besides the United States were present; however, the dataset only mentioned the U.S. forces.

Figure 57. Box Plot Casualties Side A (Government)



The greatest number of casualties suffered by side B was 872,400 during the Chinese Civil war between 1946 and 1950. The Chinese communists initiated war against the Chinese government; the communists were ultimately successful in establishing their government. The fewest casualties on side B (insurgents) was 40, during the Greek Independence conflict, fought between France and Greece in 1827.

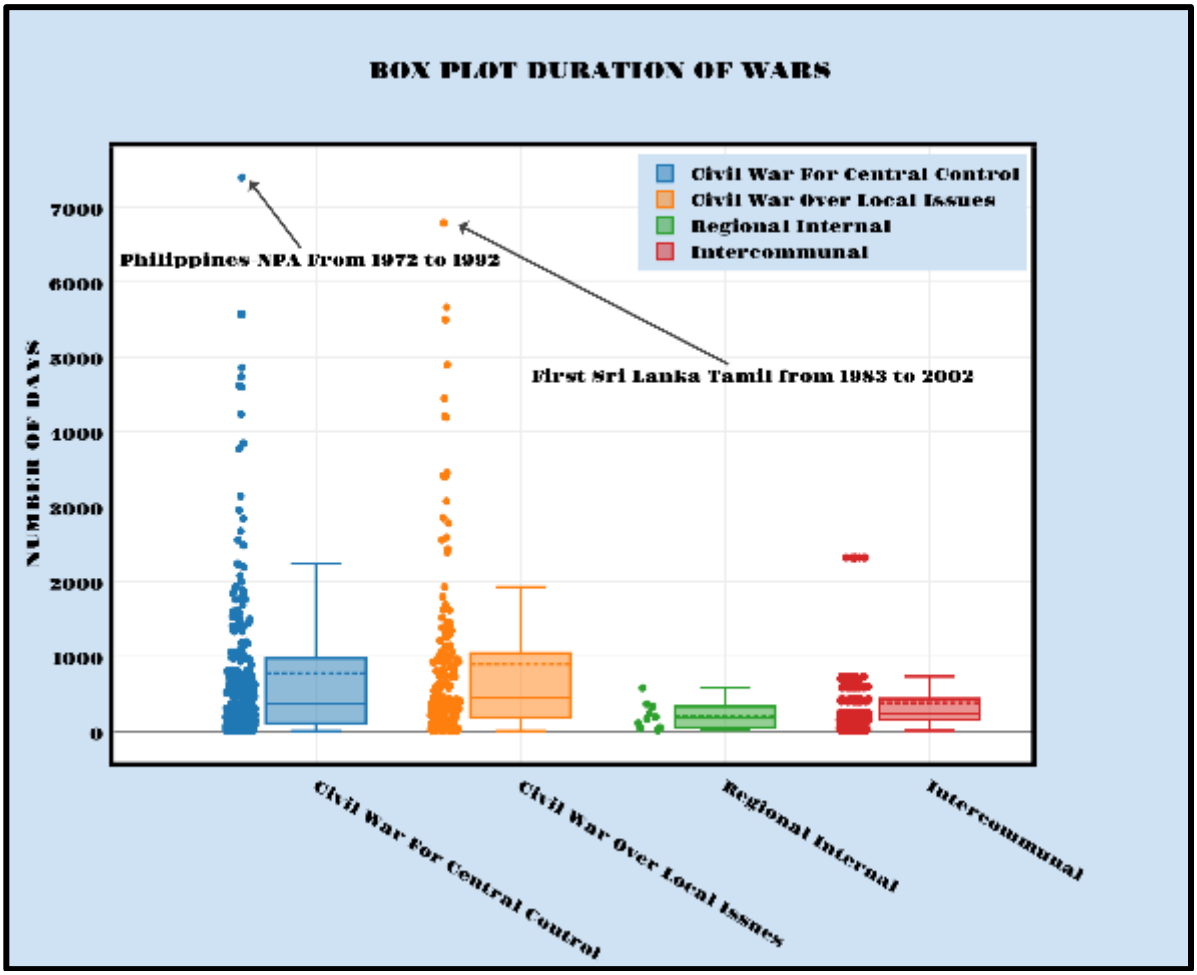
Figure 58. Box Plot Casualties Side B



4. Duration of Different Wars

The COW's intra-state war dataset consists of 442 wars. There are some limitations in modeling the wars since some wars do not have start or end dates. The Second Bosnia, Bolivian Perez Rebellion, and Yellow Cliff Revolt are the wars that are not included in analysis since they occurred during one year, but their start date and start month is unknown. There other wars in the dataset that span over more than one year, and information regarding start date, start month, end date, and end month is not available. These are estimated by using fixed start and end dates. Figure 59 displays the duration of different wars; it appears that *civil war for central control* and *civil war over local* issues are longer in duration than regional internal and inter-communal wars. The regional and inter-communal wars are spread over a relatively short span of time.

Figure 59. Duration of Wars



The summary statistics reveal different means, standard deviations, minimums, and maximums for all four war types. The mean days for civil war for central control, civil war over local issues, regional internal, and inter-communal war is 769.06, 870.94, 225.54, and 379.05, respectively. The standard deviation in days for civil war for central control, civil war over local issues, regional internal, and inter-communal war is 1,084.044; 1,177.82; 179.71; and 508.60.

5. War Duration and Number of Casualties

Intuitively, there should be more casualties if a war is extended over a greater number of days. The actual data proves this false. In the cases in which the casualties of

only one side is known, only those casualties are considered. There are certain wars for which casualties on both sides are unknown; these battles are not included in the analysis.

Figure 60. Duration of Wars and Casualties

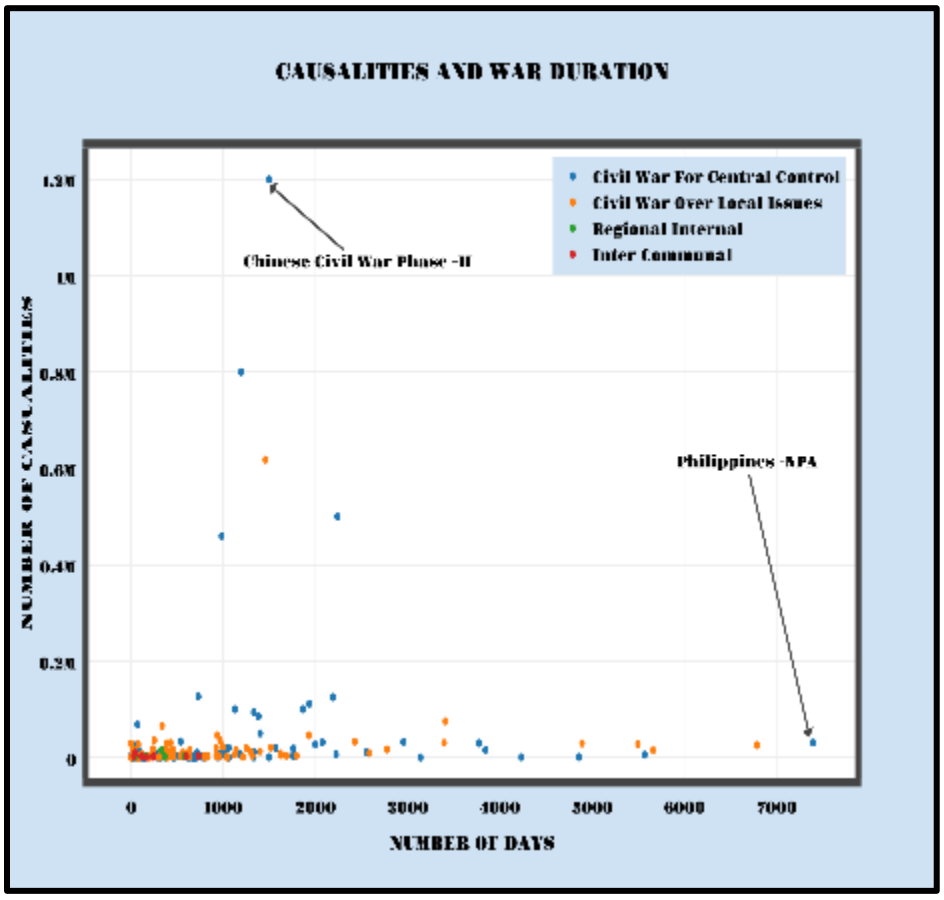


Figure 60 indicates that the majority of the conflicts took place between zero and 2,000 days, which is the equivalent of 5.64 years. The greater number of casualties occurred from the Chinese Civil War, which are within the 1,000 and 2,000 days range. Another significant factor is that regional and inter-communal wars are relatively short in duration, typically within a range of zero and 1,000 days. For example, wars with long durations such as the Philippines-NPA do not significantly reflect more casualties.

C. STATISTICAL ANALYSIS OF COW NON-STATE WAR

The Correlates of War project for non-state war includes conflict between non-state entities. According to COW, these entities include *non-state autonomous parties* that do not fulfill the criteria of *inter-state systems*, *non-territorial entities* (NTEs), or *non-state armed groups* (NSAs).[124] The COW project for non-state war includes two types of wars. The first type is *war between non-state entities*; those take place in a non-state territory. The second type is war between *non-state armed groups (NSAs)*; those take place across state borders. The dataset contains 62 *non-state wars,* out of which 61 wars are of the first type, and one war is of the second type.[125] The first war in the dataset is the First Maori Tribal war, which took place between 1818 and 1824; the last war is the Hemda-Lendu conflict from 1999 and 2005.

1. Types of War and Different Regions

The majority of wars in the dataset took place in Asia and Africa. Europe is unique in that no-war of this category took place in Europe. One of the reasons for this may be the strong nation-state system that exists in Europe. Table 60 gives a detailed summary of *non-state wars* by type in different regions.

[124] Meredith Reid Sarkees, "Non-State Wars (Version 4.0)," *Definitions and Variables.Correlates of War Project* (2010b).

[125] Ibid.

Table 60. Types of Wars in COW Non-State War Dataset

TYPOLOGIES OF WAR							
Type of War	W. Hemisphere	Europe	Africa	Middle East	Asia	Oceania	Total
War between Non-State entities in Non-State territory	15	-	18	5	19	4	61
War between NSAs across State border	-	-	1	-	-	-	1
Total	15	-	19	5	19	4	62

2. Outcome of Wars

In this dataset there is no clear difference between side A and side B since mostly both sides are non-state. The dataset consists of 62 wars, out of which 28 wars were won by side A and 17 by side B. As far as the latter wars are concerned, two resulted in compromise, eight transformed into another type of conflict, five resulted in stalemates, and two continue as conflicts below the war level. From a military perspective, the outcome of war does not provide detailed information regarding attacker, defender, or winner as was given in the CDB90G dataset. However, it gives information regarding stalemate and transformation of conflict into other form, which is also useful for military leaders. Table 63 shows some detail of the different wars.

Table 61. Types of Wars and Outcome

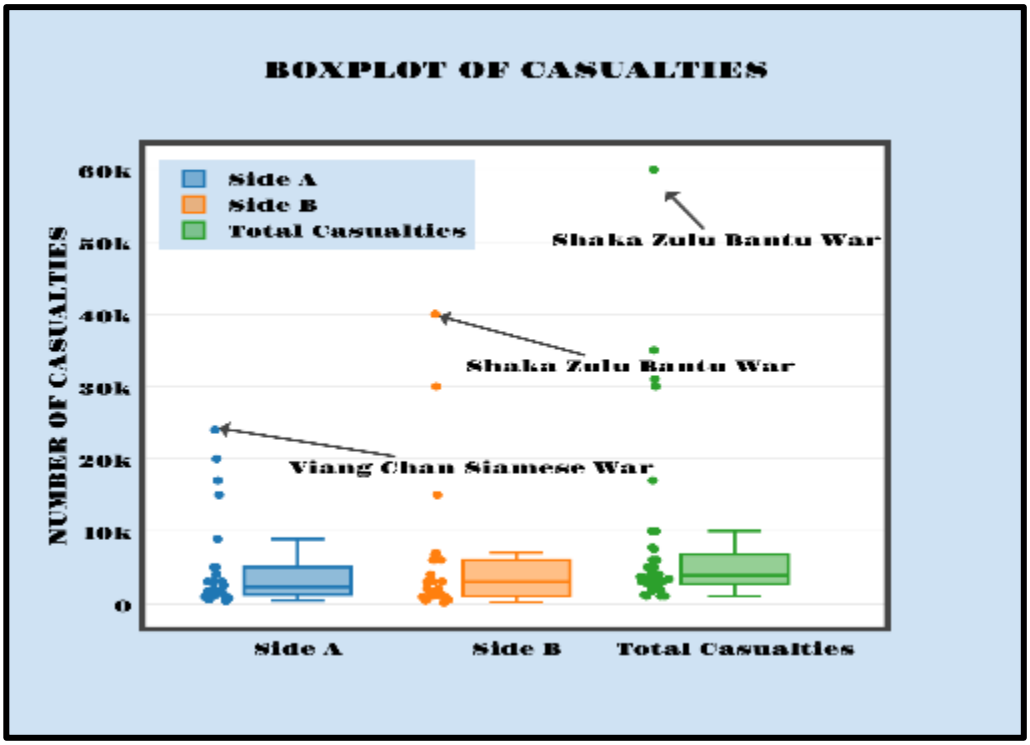
OUTCOME OF NON-STATE WARS			
Type of War	War between Non-State entities in Non-State territory	War between NSAs across State border	Total
Side A Wins	28	-	28
Side B Wins	17	-	17
Compromise	2	-	2
Transformed	8	-	8
Stalemate	4	1	5
Conflict Continues Below War level	2	-	2
Total	61	1	62

The Mexico-Yaqui Indian and Dogra-Tibet were two conflicts that resulted in compromise. The First Haiti-Santo Domingo, Han-Nien, First Boer-Basuto, Dhofar Rebellion Phase-1, and Hemda-Lendu wars resulted in stalemates. The First Australian Aboriginal and Second Australian Aboriginal wars are the conflicts that continued below war level.

3. Number of Casualties

There are 26 wars in the dataset in which total numbers of casualties are unknown. There are 14 wars in which casualties of neither side are known; for these wars, one-sided casualties make up the total. The greatest number of casualties for side A took place in the Viang Chan–Siamese conflict, for a total of 24,000. The greatest number of casualties for side B is 40,000 in the Shaka Zulu-Bantu war.

Figure 61. Box Plot Casualties of Non-State Wars

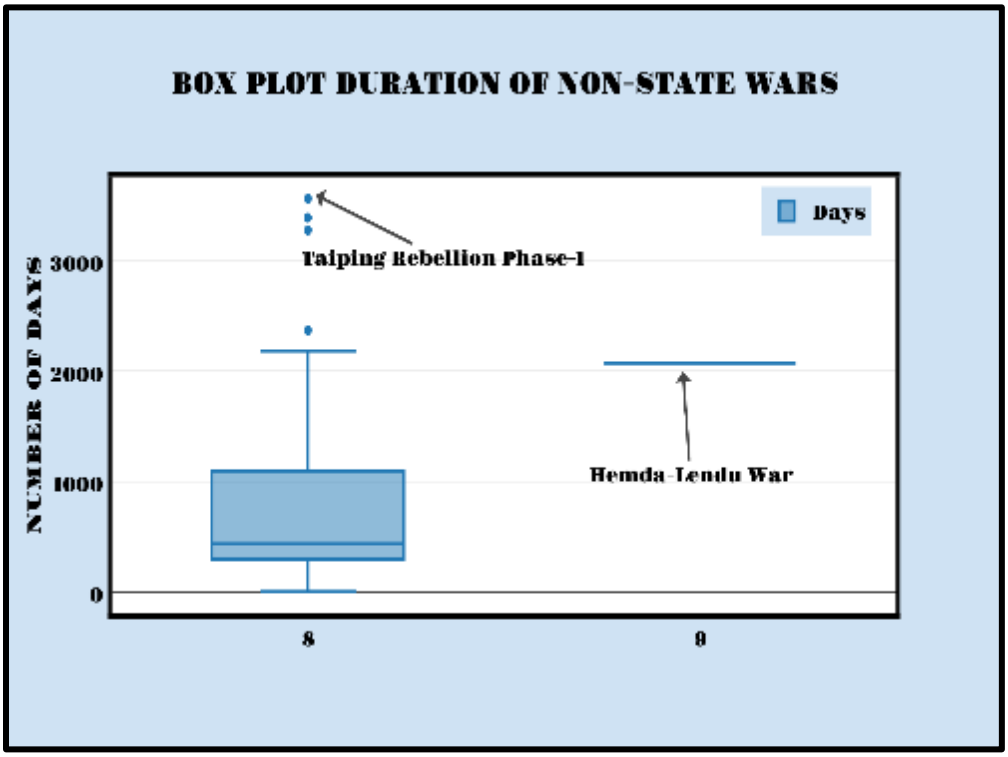


4. Duration of Wars

The analysis of the duration of non-state wars has been carried out in the same way as intra-state war. For purposes of analysis, three wars have not been considered because their start and end dates were unknown, and all three wars lasted less than one year. These wars include the Kuch and Khoja Uprising, First Boer-Basuto conflict, and Rabih Zubayr Bornu conflict, which occurred during years 1857, 1858, and 1893, respectively. For wars in which either the start date or end date was missing, that date has been approximated to the first day of the nearest month. In cases where a war spans more than one year and where the start and end date is not provided with information about exact day or month, that the missing information has been approximated by using the complete year. For example, in the case of the Burma Assam War, we only have the year when the war commenced (1819) and year when the war ended (1822). The other information, such as start date/month and end date/month, is missing. In order to determine the number of days for these kinds of cases, the number of years is used (i.e.,

three years or 1,092 days). There are two wars that span less than one day, which are the First Zulu Internecine and Nejd-Hejaz wars. There is only one war between non-state armed groups (NSAs) across state borders, which continued for 2,070 days.

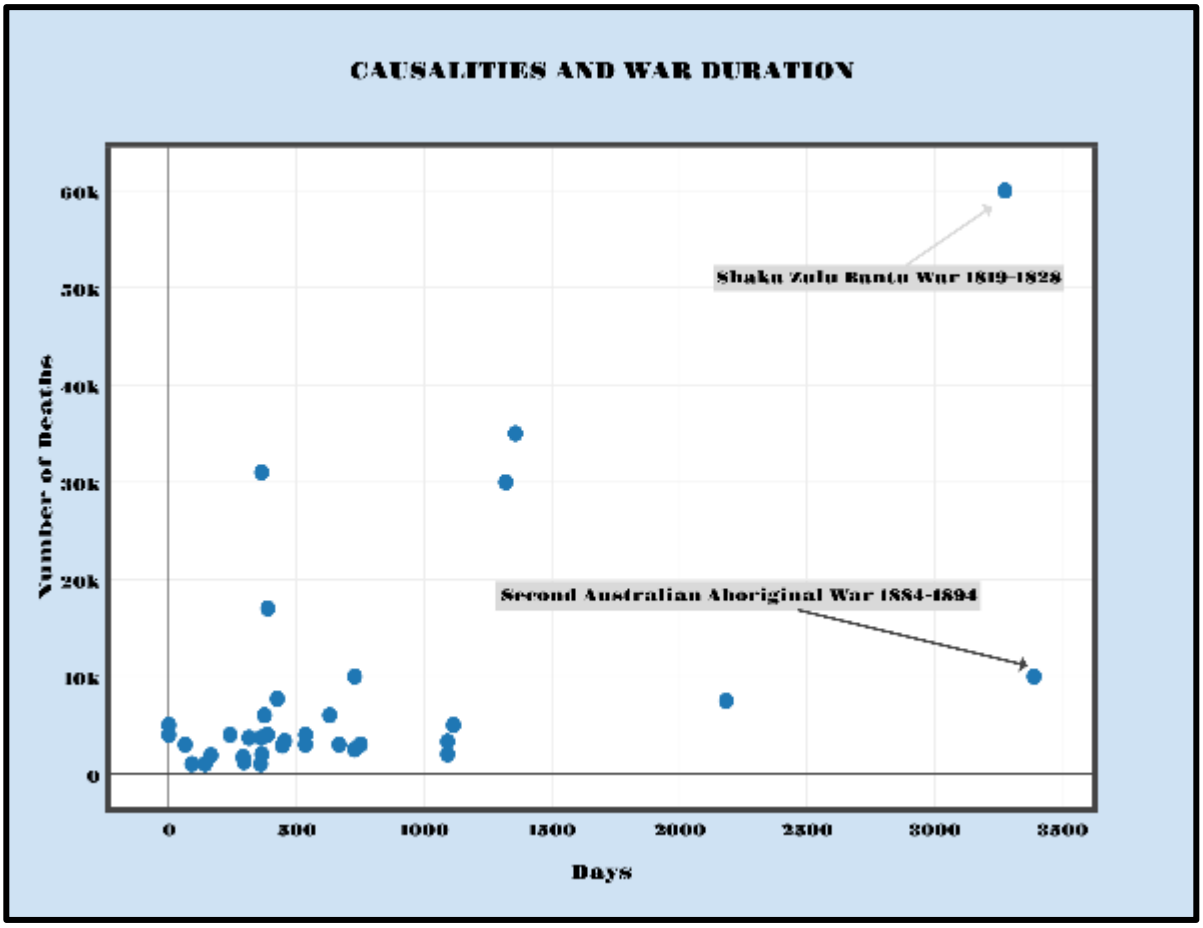
Figure 62. Box Plot Duration of Non-State Wars



5. Duration of Different Wars and Number of Casualties

Duration of war and the number of casualties seem to be directly proportional. This was verified by analyzing the duration of 35 wars against the total number of casualties. There are 24 wars in which the casualties of both sides is unknown, and three wars for which the duration is unknown. These wars have been excluded from our analysis.

Figure 63. Casualties and War Duration



The mean for casualties is 8,269.6 persons, with a standard deviation of 12,369.38. The mean duration for wars is 712.1 days, which is approximately 1.95 years. The longest war in the dataset is the Taiping Rebellion, which is not included in the analysis of casualties and war duration since the number of casualties for it is not available. The second longest war in the dataset is the Second Australian Aboriginal war, and the number of casualties in this war are slightly greater than the casualty mean for the dataset.

D. STATISTICAL ANALYSIS OF UPPSALA CONFLICT DATA PROGRAM

The Uppsala Conflict Data Program (UCDP) is a project of the Department of Peace and Conflict Research, Uppsala University, of Sweden. The project has a comprehensive database covering various aspects of armed violence since 1946. The

UCDP is considered as one of the most accurate and well-documented data sources on global armed conflict. The database of UCDP is regularly updated to incorporate ongoing conflicts. The UCDP consists of various datasets, which are geo-referenced,[126] external support,[127] actor, armed conflict,[128] monadic conflict onset and incidence, dyadic,[129] conflict termination,[130] peace agreement,[131] conflict database categorical variables,[132] non-state-conflict,[133] one-sided violence,[134] intra-state low intensity conflict (MILC)[135] and managing intra-state conflict (MIC) Africa.[136] However, for the purpose of this research, non-state conflict and one-sided violence are analyzed since in this chapter we are focusing primarily on irregular warfare.

E. STATISTICAL ANALYSIS OF UPPSALA NON-STATE CONFLICT DATASET

The non-state conflict dataset is developed by Uppsala University in collaboration with Simon Fraser University, Vancouver, Canada. The project defines non-state conflict as*: "the use of armed force between two organized armed groups, neither of which is the*

126 Clionadh Raleigh, Linke Andrew, Hegre Havard and Karlsen Joakim, "Introducing ACLED: An Armed Conflict Location and Event Dataset Special Data Feature," *Journal of Peace Research* 47, no. 5 (2010): 651–660.

127 Croicu, Cătălin Mihai, Högbladh Stina, Pettersson Therése, and Themnér Lotta, "UCDP External Support Project Primary Warring Party Dataset Codebook" (2011): 1–2011.

128 Harbom Lotta, Högbladh Stina, and Wallensteen Peter, "Armed Conflict and Peace Agreements," *Journal of Peace Research* 43, no. 5 (2006): 617–631.

129 Harbom, Lotta. "UCDP Dyadic Dataset Codebook," *Uppsala Conflict Data* (2009).

130 J Joakim Kreutz, "How and When Armed Conflicts End: Introducing the UCDP Conflict Termination Dataset," *Journal of Peace Research* 47, no. 2 (2010): 243–250.

131 Lotta Themnér and Wallensteen Peter, "Armed Conflicts, 1946–2011," *Journal of Peace Research* 49, no. 4 (2012): 565–575.

132 Uppsala Conflict Data Program, Uppsala University. "UCDP Database Categorical Variables 1989–2008" (1989-2008).

133 Ralph Sundberg, Eck Kristine, and Kreutz Joakim, "Introducing the UCDP Non-State Conflict Dataset," *Journal of Peace Research* 49, no. 2 (2012): 351–362.

134Ralph Sundberg, "Revisiting One-Sided Violence–a Global and Regional Analysis," *States in Armed Conflict, Ed.Lotta Harbom and Ralph Sundberg.Uppsala: Universitetstryckeriet* (2009).

135 Erik Melander, Frida Möller and Magnus Öberg, "Managing Intrastate Low-Intensity Armed Conflict 1993–2004: A New Dataset," *International Interactions* 35, no. 1 (2009), 58–85.

136 Erik Melander and Uexkull von Nina, "Sustained Third Party Engagement and Conflict Termination: An Introduction of the UCDP Managing Intrastate Conflict (MIC) Dataset."2011.

government of a state, which results in at least 25 battle-related deaths in a year."[137] According to the project, there are two different types of groups: formally organized and informally organized groups.

The formally organized groups, according to the project, are the ones which consist of *"any non-governmental group of people having announced a name for their group and using armed forces against another similarly formally organized group."*[138] The project also defines an informally organized group as "*any group without an announced name, but who uses armed force against another similarly organized group, where there is a clear pattern of violent incidents that are connected and in which both groups use armed force against the other."* The dataset contains information on the type/ level of organizations carrying out actions, belligerents, start date, end date, number of casualties, and region. There are a total of 784 conflicts in the dataset between 1989 and 2013.[139]

1. Type of Organizations and Conflicts in Different Regions

The Uppsala Non-State dataset consists of three different types of organizations that have been involved in conflicts in five different regions.[140] According to the UCDP Non-State Conflict Codebook version 2.5-2014, the organization level one is *formally organized groups*; these include rebel groups and other organizations that are highly organized to participate in state-based armed conflict category; for example, Ansar-ul-Islam, Afganistan.[141] The organizational level two is an *informally organized group*; these are groups associated with political parties. According to the codebook of the dataset in this category there is also fighting between groups composed of supporters of other organizations, such as the supporters of the Al-Ahly football team fighting against the supporters of the Al-Masry football team in Egypt 2012. The organization level three

---

137 Ralph Sundberg, Eck Kristine, and Kreutz Joakim, "Introducing the UCDP Non-State Conflict Dataset," *Journal of Peace Research* 49, no. 2 (2012): 351–362.

138 Ibid.

139 Ibid.

140 Ibid.

141 Ibid.

is *informally organized group*; these are groups that are not organized for fighting but organize themselves for fighting on a required basis; for example, Hindus of India fighting against Muslims of India. These three levels of organization reflect the expertise of various groups, and we use information from the dataset to evaluate which groups are mostly involved in conflicts. The maximum numbers of conflicts are carried out by organizations having skill level three, or in other words, *informally organized groups*. The maximum numbers of conflicts are carried in Africa with 520 conflicts. These results are quite similar to the results of the ACLED dataset in which most of the non-state conflicts were in Africa. Table 62 gives complete details of the conflicts in different regions with different types of organizations.

Table 62. Uppsala Non-State Conflict Regions

NON-STATE CONFLICT REGIONS						
Type of Organization	Europe	Middle East	Asia	Africa	America	Total
Organization Level 1 (Formally Organized Groups)	9	43	78	153	63	346
Organization Level 2 (Informally Organized Groups)	-	2	5	21	2	30
Organization Level 3 (Informally Organized Groups)	4	6	48	346	4	408
Total	13	51	131	520	69	784

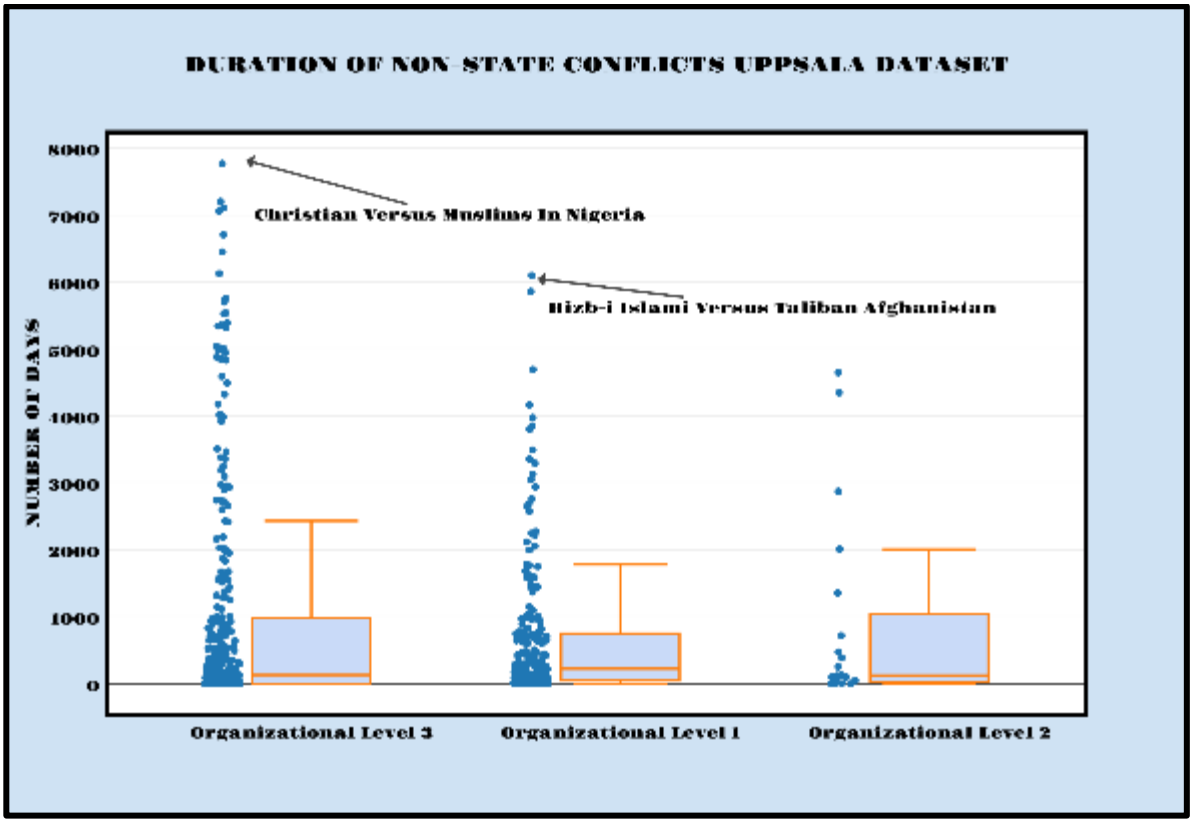
2. Duration of Conflicts

The dataset contains 784 battles, but start and end dates for all conflicts are not given. There are 231 conflicts with no end date and three conflicts with an end date before the start date. The conflicts with end date before start date are Hizb-i-Wahdat versus Hizb-i-Wahdat-Akbari faction, Taliban-Ali Dad versus Taliban-Mola Khel faction, and Forces Républicaines de Côte d'Ivoire (Republican Forces of Ivory Coast),

abbreviated as FRCI versus Movement for the Liberation of Western Ivory Coast, abbreviated as MILOCI.

There are 123 conflicts that terminate in one day. The longest conflict in the dataset is between Christians and Muslims of Nigeria, which lasted for 7,770 days and was carried out by organization of level three (*informally organized group*). The longest conflict for *formally organized group* was Hizb-i-Islami versus Taliban that continued for 6,095 days. The mean duration of conflict for organization levels 1, 2, and 3 is 667.38, 879.9, and 975.17 days with standard deviations of 1,055.12; 145.82; and 1,678.58; respectively. The medians for duration of conflict are 223, 123.5, and 130 days. Figure 64 shows box plots of duration of conflicts based on three different types of organizations; most wars are below the threshold of 1,000 days.

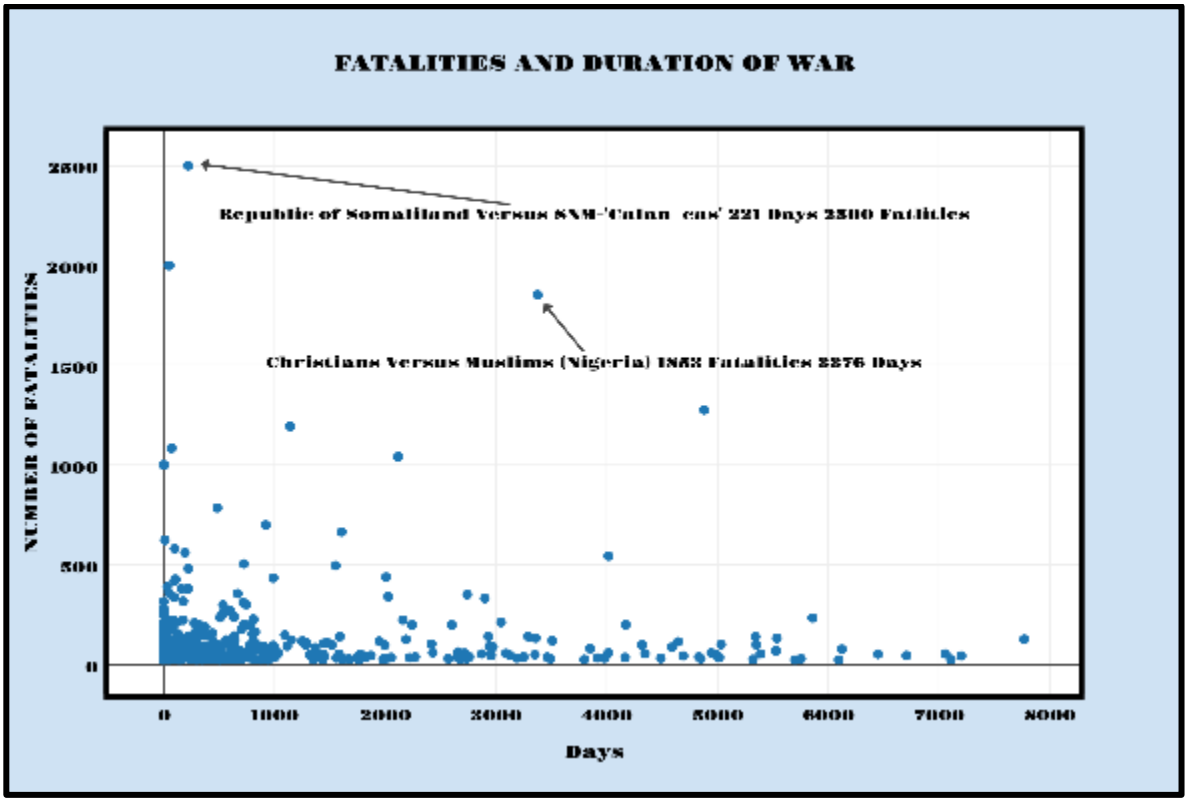
Figure 64. Duration of Non-State Conflicts Uppsala Dataset



3. Fatalities and Duration of Conflicts

Most conflicts in the dataset are within the range of zero to 1,000 days. Maximum casualties are also within the same range. The maximum number of casualties in the dataset took place in the conflict between the Republic of Somaliland versus SNM—"Calan-cas" which was carried out for 221 days and resulted in approximately 2,500 fatalities. The longest conflict in terms of duration is the conflict between Christians and Muslims of Nigeria; however, the number of casualties during the conflict were low in comparison to the Somaliland war. The mean for casualties in the dataset is 118.07 with a standard deviation of 227.61. The maximum casualties in the dataset are 2,500 and minimum casualties are 25. Figure 65 demonstrates the relationship between duration and casualties.

Figure 65. Fatalities and Duration of Conflicts



F. STATISTICAL ANALYSIS OF UPPSALA ONE-SIDED VIOLENCE DATASET

The one-sided violence dataset is a collaborative project of the Uppsala Conflict Data Program (UCDP) and Human Security Report Project, Simon Fraser University, Vancouver, Canada. According to the project, a one-sided violence is defined as *"use of armed force by the government of a state or formally organized group against civilians, which result in at least 25 deaths. This does not include extrajudicial killing."*[142] The dataset defines state as *"an international sovereign government controlling a specified territory or an internationally unrecognized government controlling a specified territory whose sovereignty is not disputed by another internationally recognized sovereign government previously controlling the same territory."*[143] The dataset has 12 variables with 795 entries between 1989 and 2013. It covers five different regions.

1. Type of Actors and One-Sided Violence in Different Regions

The one-sided dataset has two types of actors committing violence: non-government and government. Table 63 shows the number of events by both categories of the actors in different regions. The maximum number of one-sided violence has been carried out in Africa by both government and non-government. Asia ranks second after Africa with 244 events of one-sided violence.

Table 63. Uppsala One-Sided Violence Conflict Regions

ONE SIDEDVIOLENCE CONFLICT REGIONS						
Type of Organization	Europe	Middle East	Asia	Africa	America	Total
Non-Government	19	47	181	238	41	526
Government	14	25	63	151	16	269
Total	33	72	244	389	57	795

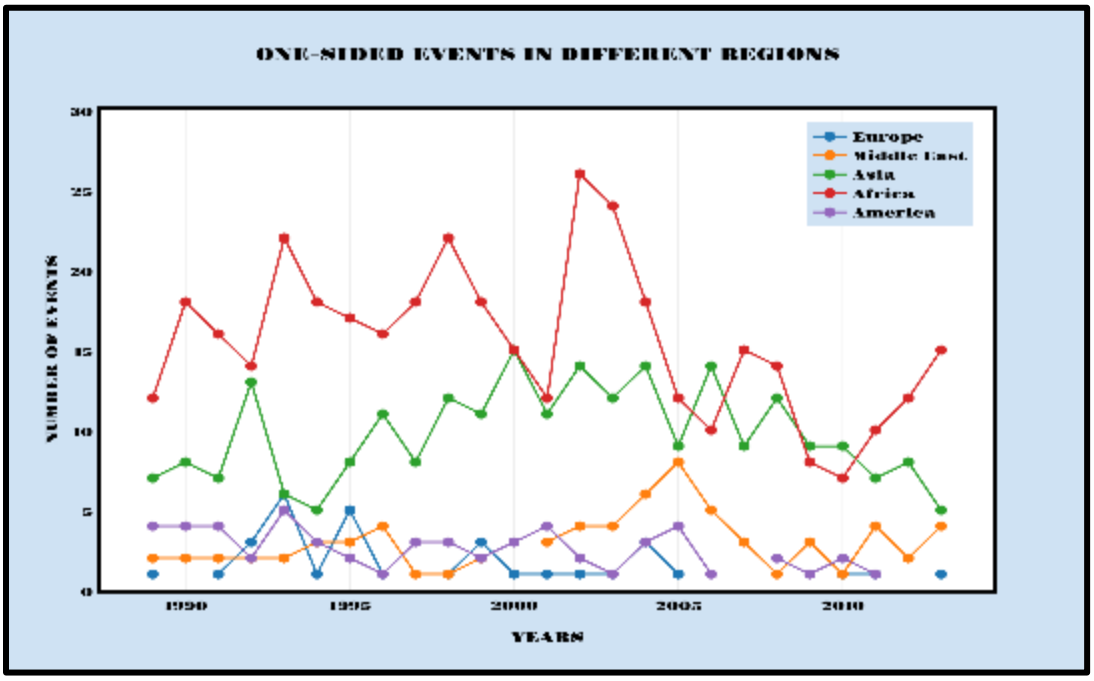
[142]Kristine, Eck, and Hultman Lisa, "One-Sided Violence Against Civilians in War Insights from New Fatality Data," *Journal of Peace Research* 44, no. 2 (2007): 233–246.

[143] Ibid.

2. One-Sided Violence During Different Years

The dataset contains events of one-sided violence between 1989 and 2013. The dataset provides years of incidents for a particular event that occurred without giving complete temporal details. Figure 66 indicates different events according to different years and regions. The red line representing Africa clearly indicates that Africa remained on top of one-sided violence as compare to the other regions. The line also indicates that there is still an inclining trend of violence in Africa. In other regions of the world, there has been a decrease in violence. The most one-sided violence witnessed by Africa is during 2002. Figure 66 also shows that America and Europe have less one-sided violence than Africa and Asia do.

Figure 66. One-Sided Violence in Different Years

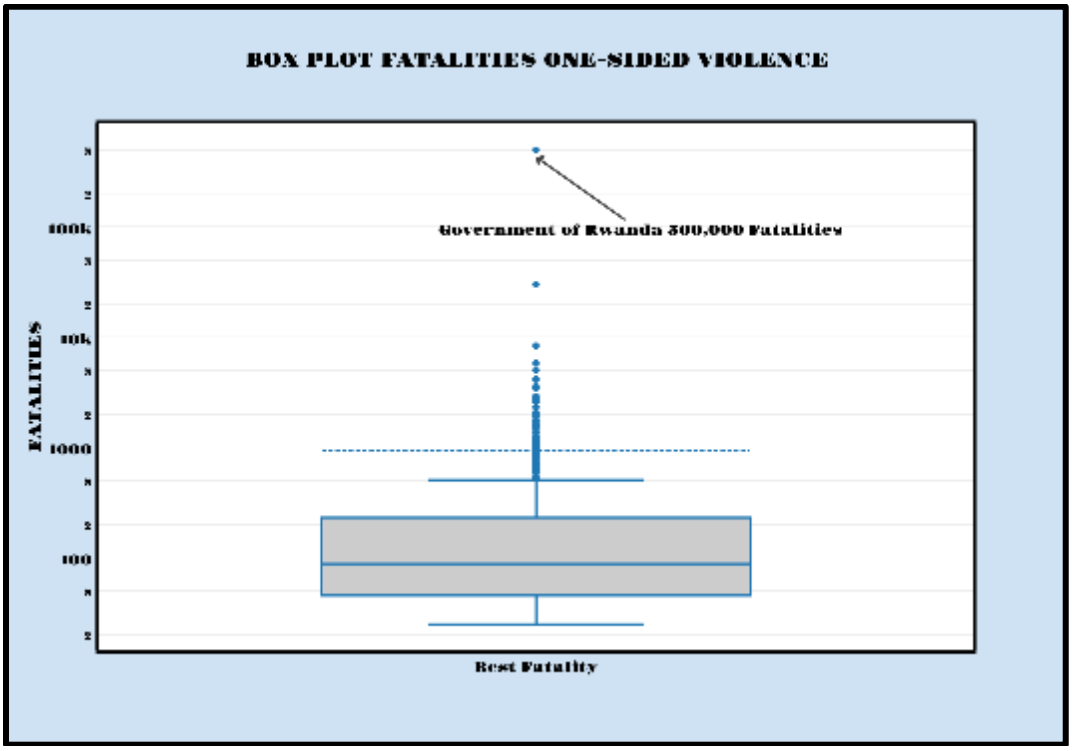


3. Fatalities from One-Sided Violence

The dataset has recorded fatalities in three different categories: best, low, and high. For the purpose of analysis we use the best category, which is the most accurate estimation of casualties. The mean fatalities for one-sided violence is 938.75, with a

standard deviation of 17,763.99. The minimum fatalities during any particular event are 25 and maximum are 500,000. Figure 67 shows a box plot of fatalities for one-sided violence.

Figure 67. Box Plot Fatalities for One-Sided Violence



THIS PAGE INTENTIONALLY LEFT BLANK

VI. CLASSIFICATION AND REGRESSION TREE MODELS

This chapter uses Tree techniques to confirm the results of the descriptive statistics covered in Chapter III and determines the important variables for winning battles in conventional warfare. In this chapter, classification models are made based on the CDB90G dataset to predict the outcome of conventional battles based on variables discussed in Chapter III. The CDB90G dataset is modeled using R.[144] Moreover, the ACLED dataset is modeled to determine casual factors for irregular conflicts using the boosted tree methodology in JMP.[145]

A. INTRODUCTION

Trees are a method used in data mining and machine learning for predictive modeling.[146] Trees are of two types: classification and regression trees. Regression trees are used when the response variable is numeric or continuous; for example, price related to different items. If the response variable is discrete or a finite set of values, classification tree methodology is used. The CDB90G dataset is examined using classification tree methodology.

Classification tree uses predictor variables to predict the classification of a response variable. The response variable can have different classes, such as *Yes or No*, *Male or Female*, *Apples or Oranges*, etc. In a classification tree, data are split to maximize dissimilarity among the resulting subsets based on the likelihood ratio of statistics and deviance measures.[147] In other words, data are to increase the homogeneity of the two sub-nodes.

144 R. Core Team, *R Language Definition* (2000).

145 JMP Pro, v10, *SAS* (2012).

146 Breiman, Leo, Jerome Friedman, Stone J Charles, and Olshen A Richard, *Classification and Regression Trees* (Boca Raton, FL: CRC Press, 1984).

147 Matthew Hansen, Dubayah R.and DeFries R., "Classification Trees: An Alternative to Traditional Land Cover Classifiers," *International Journal of Remote Sensing* 17, no. 5 (1996), 1075–1081.

To further simplify the concept of classification trees, we discuss a simple example in which we have two variables, age and sex, which predict whether a person will sign up at a gym. Our training data reflect that 85% of the persons who are above 40 years old signed up for a gym. Taking the criterion of age into account, we split the data, and age becomes a top node in the tree. According to Faraway, in classification trees a split divides the observations within a node so that the class types within each split are mostly the same.[148] The purity of a node in a classification tree can be measured by several means, which include Deviance, Entropy and the Gini index.[149] All these measures will be minimized when node members are the same. Tree models are often preferred since they are simple and provide a clear picture of the data. They are considered a good tool for initial data inspection and provide highly intuitive insight into interactions between various variables.[150] Moreover, tree models are robust toward missing predictor variables and tend to accurately predict even when the data have high variability.

In the succeeding paragraphs, the CDB90G dataset is modeled into three different types of models. The first type consists of objective variables, as selected in Chapter III. The second type includes both objective and relative variables. The third model also includes conditions of weather and terrain, along with the objective and relative variables of Models 1 and 2.

B. MODEL 1 (OBJECTIVE VARIABLES)

The first types of models are built based on only objective variables. These objective variables include force size, tank, artillery, initial force, cavalry, and close air support ratios, along with attacker primary tactical scheme and defender primary posture. Models are built in the chronological order in which battles took place. This gives us trends of changing variables in winning the battles.

---

148 Julian J. Faraway, *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*CRC press, 2005).

149 Ibid.

150 Michael J. Crawley, *The R Book* (Hoboken, NJ: John Wiley & Sons, 2012).

1. Model 1.1 (1600–1697)

Model 1.1 consists of 47 battles of the seventeenth century, including the era from 1600 to 1697. The complete era has been divided into a test and training set. The training set consists of 27 battles from 1600–1650, and the test set consists of 20 battles from 1651–1697. In this dataset, cavalry emerges as one of the most important winning factors. The misclassification rate on the training dataset is 0.259, and on the test dataset it is 0.20. The misclassification rate on the test dataset is lower than on the training dataset, which indicates that model is predicting with reasonable accuracy.

Figure 68. First Split of Tree Model 1.1: Battles between 1600 and 1697

Figure 68 represents that the tree has accurately predicted 15 times when the attacker has won the battle. There are only four times on this branch when the model was unable to accurately predict the outcome. In the case of the defender, the model has predicted accurately five times. There are three times on this branch when the model was unable to predict the accurate outcome for defenders. This model is very accurate, but it is primarily based on one variable. For the purpose of military analysis, we need to grow the tree bigger in order to predict other important variables for attaining victory.

Figure 69. Tree Model 1.1a: Battles between 1600 and 1697

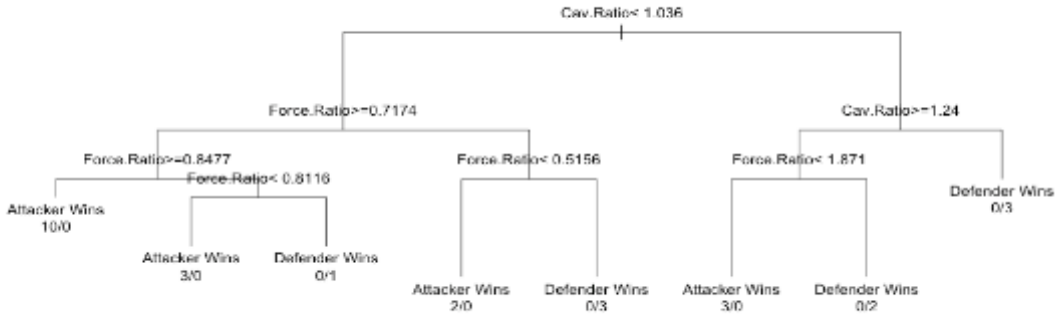
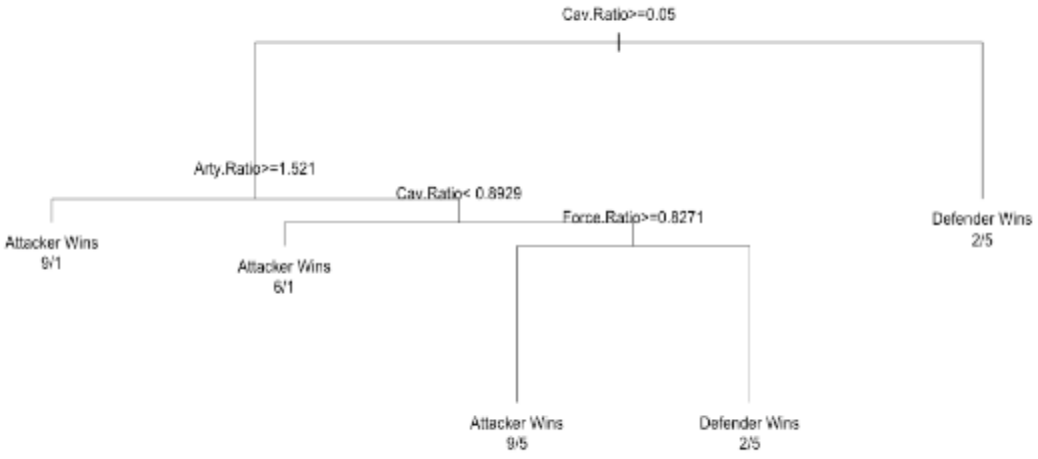


Figure 69 represents a bigger tree in order to capture other variables required to win battles during the seventeenth century. This model shows that force ratio is another important variable for winning battles in addition to cavalry. The misclassification rate for the training set in this case is zero and for the test set is 0.3, which indicates that this model is likely over-fitted. However, both models give statistical insight on winning battles in the seventeenth century.

2. Model 1.2 (1704–1799)

Model 1.2 is based on battles of the eighteenth century. The dataset comprises 63 battles between 1704 and 1799. This dataset shows that cavalry and artillery appear as important variables in winning battles. Cavalry holds its position as the most important battle-winning factor. The misclassification rate for the training dataset is 0.244, and for the test dataset it is 0.333.

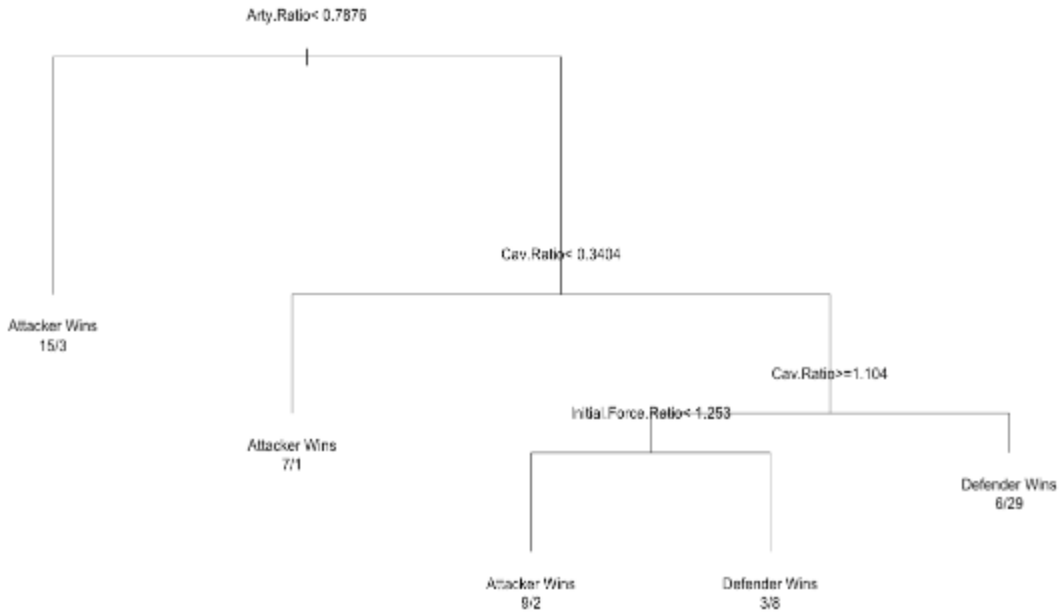
Figure 70. Tree Model 1.2: Battles between 1704 and 1799



3. Model 1.3 (1800–1893)

Model 1.3 consists of 122 battles between 1800 and 1893. The training set comprises83 battles between 1800 and 1863. The test set consists of 39 battles from 1864–1893.The misclassification for the training set is 0.180 and for the test set is 0.48. This model is comparatively less accurate than the other two models. However, variables that appear as important during this era are artillery, cavalry, and initial force ratio.

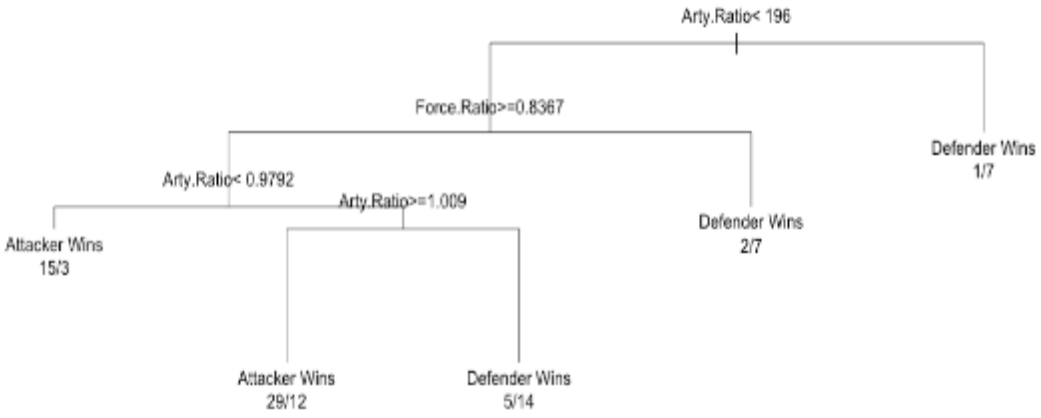
Figure 71. Tree Model 1.3: Battles between 1800 and 1893



4. Model 1.4 (1900–1920)

Model 1.4 consists of battles between 1900 and 1920. There are 125 total battles in the dataset. The training set comprises 95 battles, and the test set consists of the remaining 30 battles. The misclassification for the train dataset is 0.24 and for the test dataset is 0.19, which is lower than the training dataset. This model is accurate in prediction, and two variables that appear to be important during this time frame are artillery and force ratio.

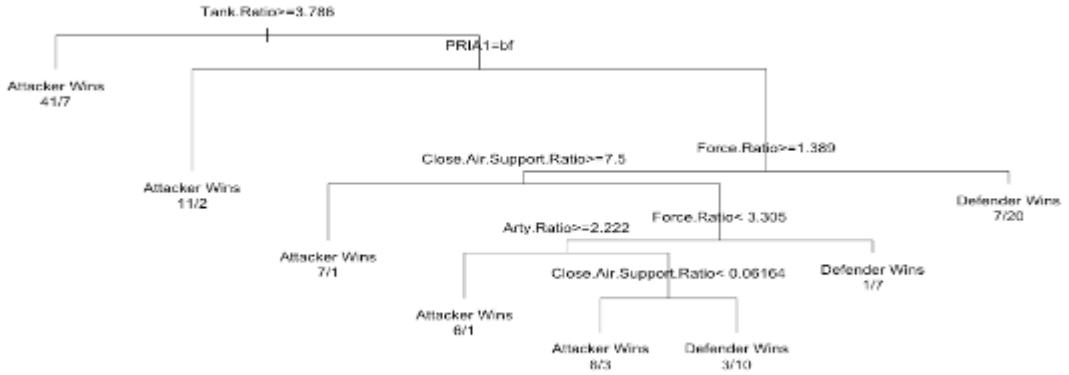
Figure 72. Tree Model 1.4: Battles between 1900 and 1920



5. Model 1.5 (1937–1945)

This model consists of battles between 1937 and 1945. There are 182 battles in total in the dataset. The misclassification rate for the training dataset is 0.187, and for the test dataset is 0.186. The tank ratio, attacker primary tactical scheme, force ratio, and close air support emerge as important variables. Despite noise in the data, this model is predicting quite accurately.

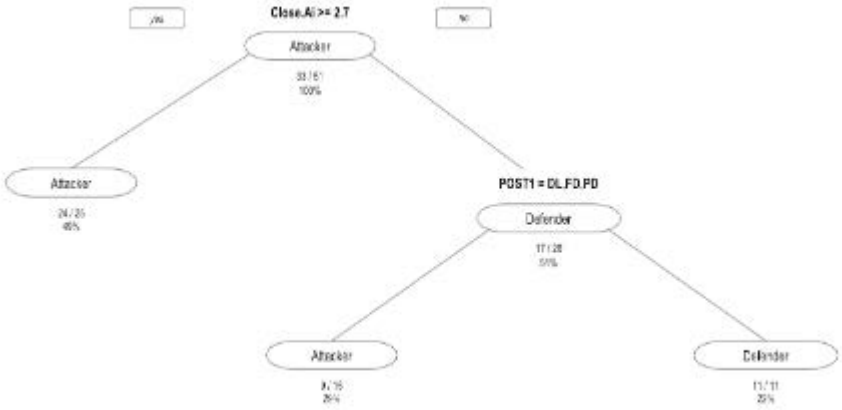
Figure 73. Tree Model 1.5: Battles between 1937 and 1945



6. Model 1.6 (1951–1982)

Model 1.6 consists of 69 battles between 1951 and 1982. These battles have a lot of variation since this was an era in which technology was rapidly changing. The two most significant variables that emerge in this model are close air support and the defender's primary posture of defense. The misclassification rate for the training dataset is 0.133 and for the test set is 0.66.

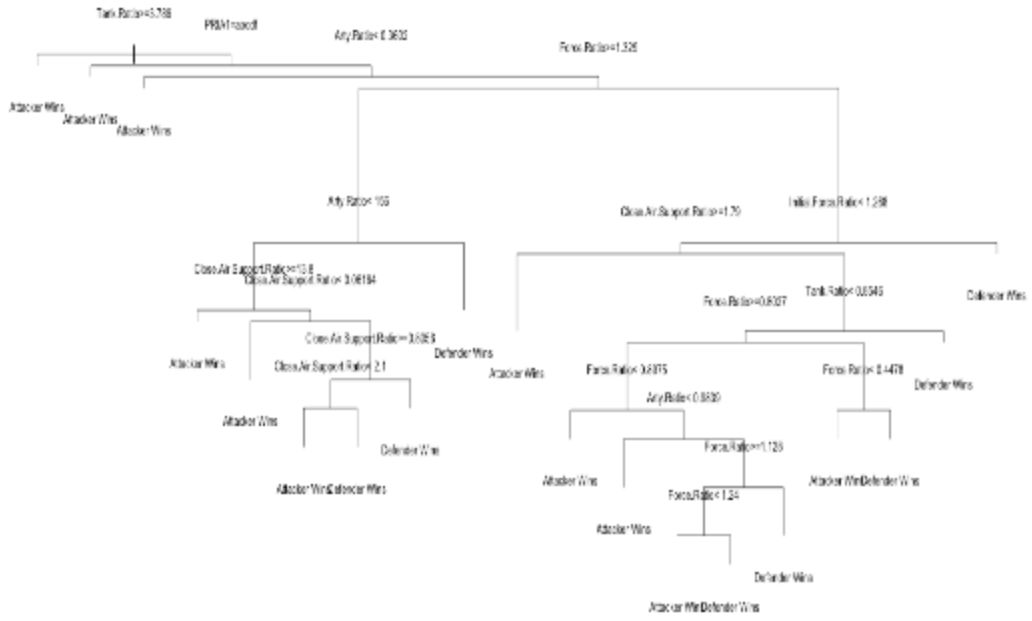
Figure 74. Tree Model 1.7: Battles between 1951 and 1982



7. Model 1.7 (1600–1982)

This model consists of a complete dataset of battles between 1600 and 1982. In this dataset there are 618 battles. The purpose of this model is to determine important variables and explore variables that are not identified in previous models. Interestingly, in this dataset tank ratio, attacker primary scheme of maneuver, artillery ratio, force ratio, initial force ratio, and close air support emerge as prominent variables. All these variables have been identified in previous models, but tank ratio emerges as the most important variable in this model. The likely reason for this is the presence of a large number of tank battles in the dataset. The misclassification rate for this model is 0.237.

Figure 75. Tree Model 1.8: Battles between 1951 and 1982



C. MODEL 2 (OBJECTIVE AND RELATIVE VARIABLES)

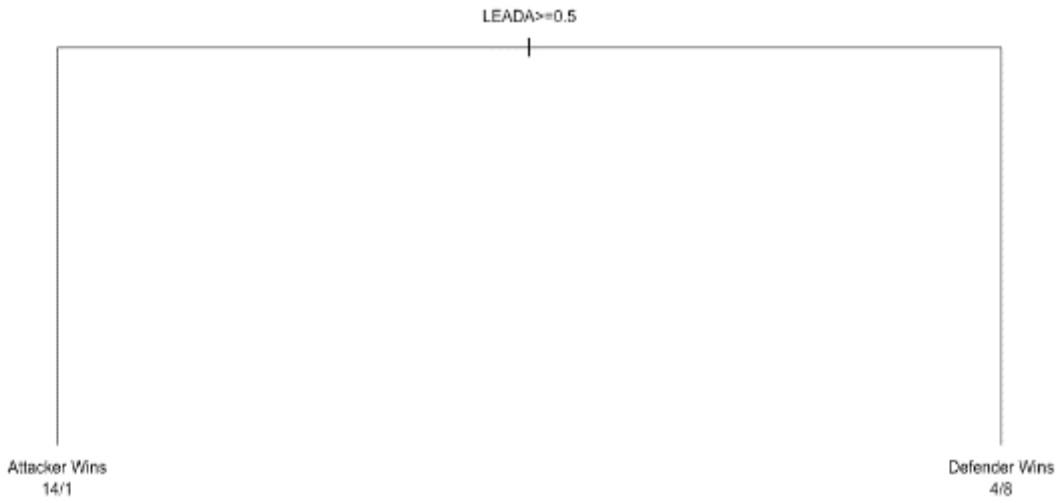
Model 2 subsets are formed using objective and relative variables. The objective variables such as ratios of force, tanks, artillery, initial force, cavalry, and close air support as well as attacker primary tactical scheme and defender primary posture are not enough to predict warfare outcomes. Warfare is a complex phenomenon and requires intrinsic analysis for determining war-winning factors and long-term trends. It involves both quantitative and qualitative analysis with insight into numbers and intangible variables, which are difficult to measure according to a specific scale. The Model 2 subsets focus on both objective and relative variables to identify factors that remain dominant during the past for winning battles.

1. Model 2.1 (1600–1697)

Model 2.1 is comprised of 47 battles between 1600 and 1697. The data are divided into two sets: a training set consisting of battles between 1600 and 1650, and the test set consisting of battles between 1651 and 1697. The training set has 27 battles and

the test set has 20 battles. The misclassification rate for the training set is 0.18 and the test dataset 0.1. This model is quite accurate in predicting the outcome of battles. The most important variable that emerges during this model is relative leadership advantage. It is also important to note that leadership judgments were made subjectively long after the battles took place. The tree based on this model accurately predicts attacker victory 14 times and defender victory eight times (see Figure 76).

Figure 76. Tree Model 2.1: Battles between 1600 and 1697



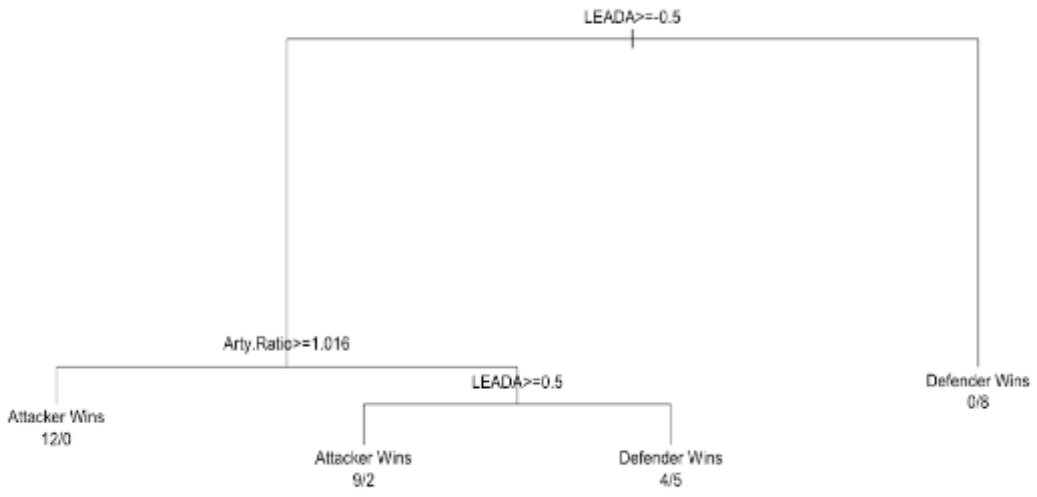
Model 1.1, which consists of only objective variables, has a higher misclassification rate than Model 2.1. Moreover, in Model 1.1 the most important split is cavalry ratio, whereas in this model the most important variable is relative leadership advantage. If we consider the time frame of this dataset and correlate historically, the variables identified by this model can be intuitively observed.

2. Model 2.2 (1704–1799)

Model 2.2 comprises 63 battles from the eighteenth century. The dataset consists of battles between 1704 and 1799. The training set has 40 and the test set has 23 battles in it. The misclassification rate for the training dataset is 0.15 and the test dataset is 0.14. In

this model, a misclassification rate is improved in comparison to the misclassification rate for Model 1.2, which consists of only objective variables for the same time frame.

Figure 77. Tree Model 2.2: Battles between 1600 and 1697

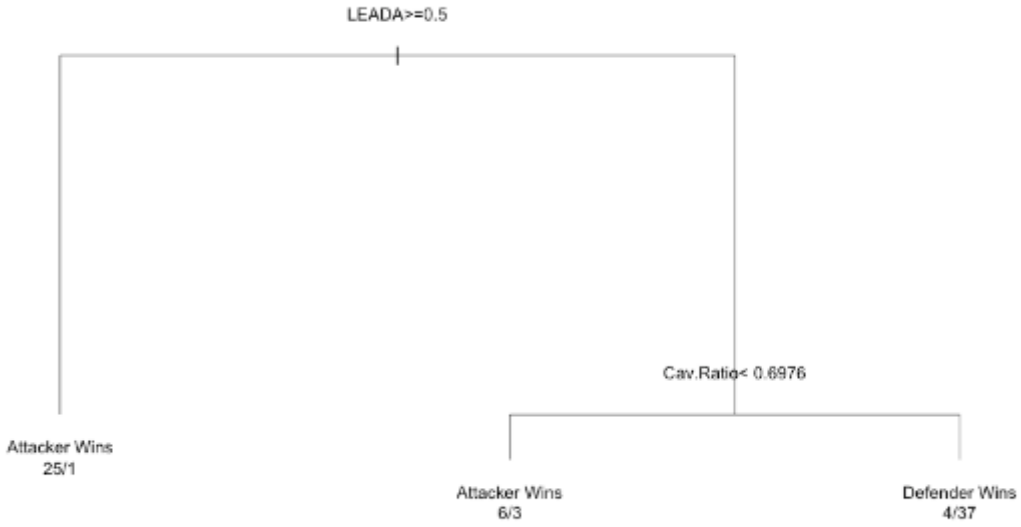


In this model, leadership and artillery ratio appear to be important variables. The primary splits for node one in this model consist of relative variables such as leadership, intelligence, combat effectiveness, and surprise. The only objective variable in the primary split for node one is force ratio. The surrogate split for node one consists of combat effectiveness and artillery ratio.

3. Model 2.3 (1800–1893)

Model 2.3 consists of battles between 1800 and 1893. In this dataset, there are a total of 122 battles. The training set consists of 76 battles and the test set consists of 46 battles. The variability in this dataset is higher than in the other model due to the wide spread of battles between 1800 and 1893. Development of new techniques and technologies during that era has also increased the variability of data. The misclassification rate for the training dataset is 0.105 and the test dataset is 0.369, which is lower than in with Model 1.3 that only includes objective variables.

Figure 78. Tree Model 2.3: Battles between 1800 and 1893

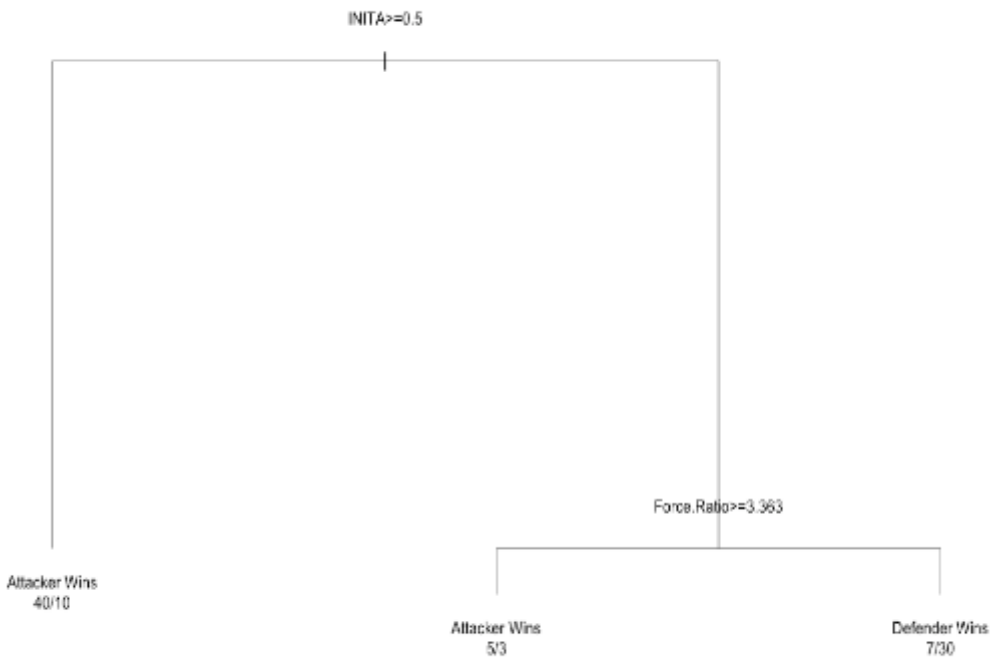


The primary splits for node one in this model are leadership advantage, initiative, combat effectiveness, artillery, and initial force ratio. The surrogate splits for node one include combat effectiveness, artillery, force, initial force, and cavalry ratio. The important variables in this model are leadership advantage, combat effectiveness, cavalry, artillery, force, and initial force ratios.

4. Model 2.4 (1900–1920)

Model 2.4 consists of battles between 1900 and 1920. There are a total of 125 battles in the dataset. The training dataset comprises 95 battles and the test dataset consists of the remaining 30 battles. The misclassification rate for the training dataset is 0.21 and for the test dataset is zero. This model is more accurate than Model 1.4 that only accounts for objective variables.

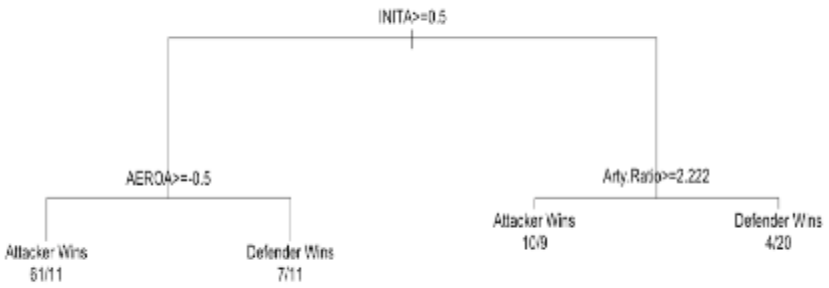
Figure 79. Tree Model 2.4: Battles between 1900 and 1920



5. Model 2.5 (1937–1945)

Model 2.5 consists of battles between 1937 and 1945. There are a total of 182 battles in the dataset. The misclassification rate for this model is higher than the rate for Model 1.5. The training dataset misclassification rate is 0.233 and the test dataset misclassification rate is 0.27. The primary splits at node one are initiative advantage, tank ratio, and relative air superiority in theatre, initial force, and force ratio. The surrogate splits are relative leadership, combat effectiveness, intelligence, momentum advantage, and tank ratio.

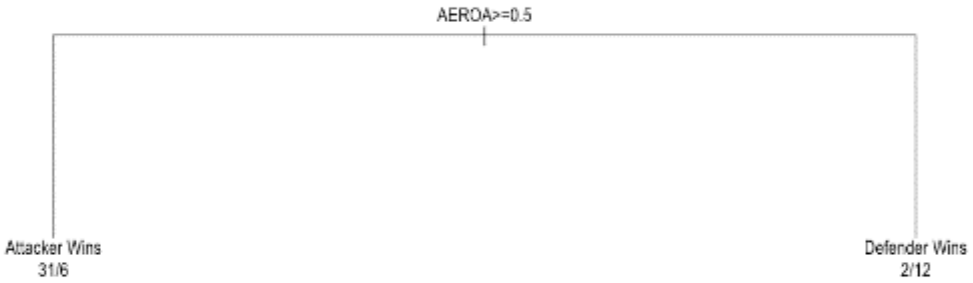
Figure 80. Tree Model 2.5: Battles between 1937 and 1945



6. Model 2.6 (1951–1982)

Model 2.6 consists of battles between 1951 and 1982. The variation in the dataset results in the highest misclassification rate for the test dataset. From 1951 to 1982 there was tremendous change in technology and warfare. The misclassification rate for the training dataset is 0.15 and the test dataset is 0.38. The primary splits for node one consist of air superiority, close air support, artillery, and attacker primary scheme. The surrogate splits consist of close air support, artillery, force, initial force, and intelligence.

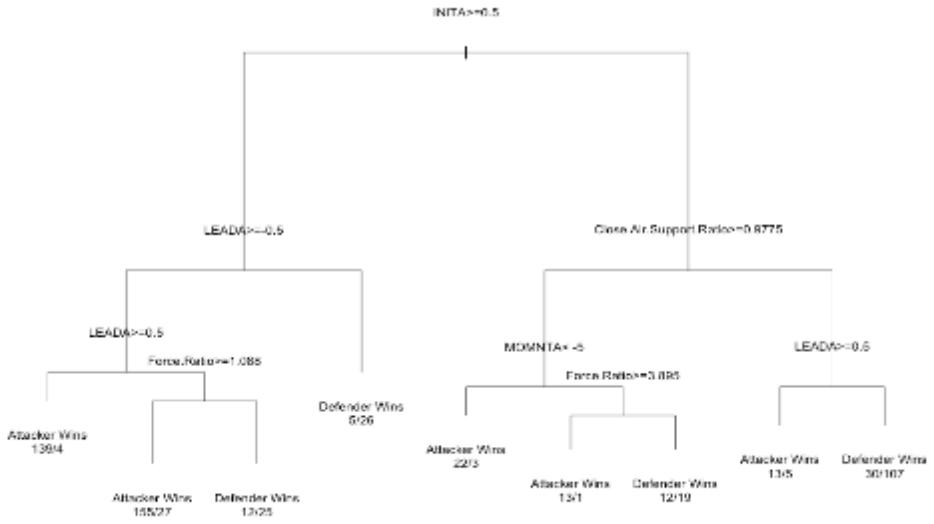
Figure 81. Tree Model 2.6: Battles between 1951 and 1982



7. Model 2.7 (1600–1982)

Model 2.7 consists of the complete dataset of battles between 1600 and 1982. In this dataset there are 618 battles. The purpose of this model is to determine important variables. The misclassification rate for this model is 0.16. The misclassification rate for this model is lower than the misclassification rate of Model 1.7, which is modeled on only objective variables.

Figure 82. Tree Model 2.7: Battles between 1600 and 1982



The important variables that appear in this model are initiative, close air support, leadership advantage, and force ratio. The primary splits for node one in this model consist of initiative, leadership, combat effectiveness, tank ratio, and air superiority. The surrogate splits consist of leadership, combat effectiveness, intelligence, momentum, and air superiority.

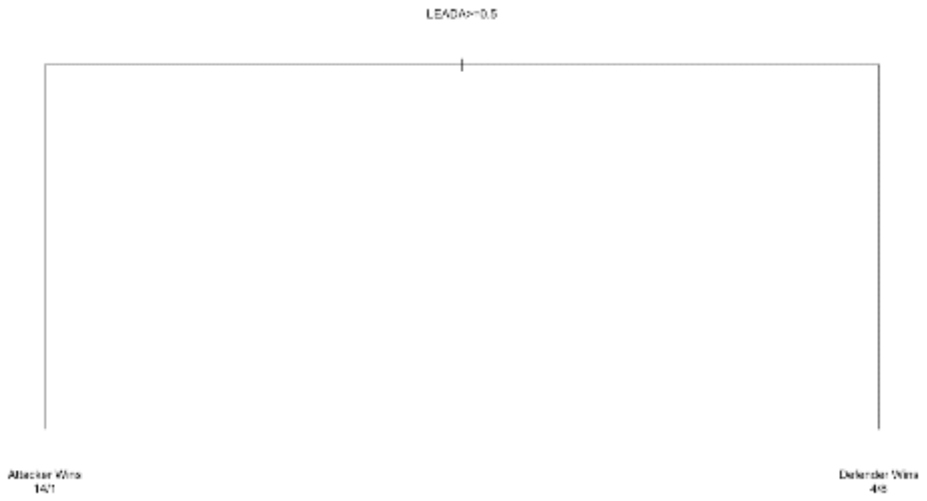
D. MODEL 3 (OBJECTIVE VARIABLES, RELATIVE VARIABLES AND CONDITIONS)

Model 3 subsets are based on objective variables and the relative variables of Model 2, along with three more relative variables, which are logistic, training, and technology advantages. Moreover, terrain and weather have also been incorporated into Model 3. The Model 3 subsets consist of up to 19 predictor variables and one response variable. The modeling is carried out based on the same time line in order to carry out analysis of additional variables in Model 3. As far as logistic, training, and technology advantages are concerned, military planners have considered them important. However, in the CDB90G dataset, information regarding these variables is noisy and very limited. In spite of this, these variables are modeled along with the other variables to determine if they have any significant impact on the outcome of battles. The descriptive statistics for these variables have been discussed in Chapter III. However, for recapitulating important information, it is essential to inform readers that in the dataset, logistic, training, and technology advantages neither favor attacker nor defender in 550, 413, and 594 battles, respectively. This means that there is very little descriptive information available on these variables. The Model 3 subsets are an extension of the two basic models.

1. Model 3.1 (1600–1697)

Model 3.1 is based on battles of the seventeenth century between 1600 and 1697. There are 47 battles in this subset. The misclassification rate for the training dataset is 0.18 and for the test dataset is 0.1. The leadership advantage appears to be an important variable in this model. In Model 1.1 and 2.1, cavalry and leadership advantage emerged as important variables.

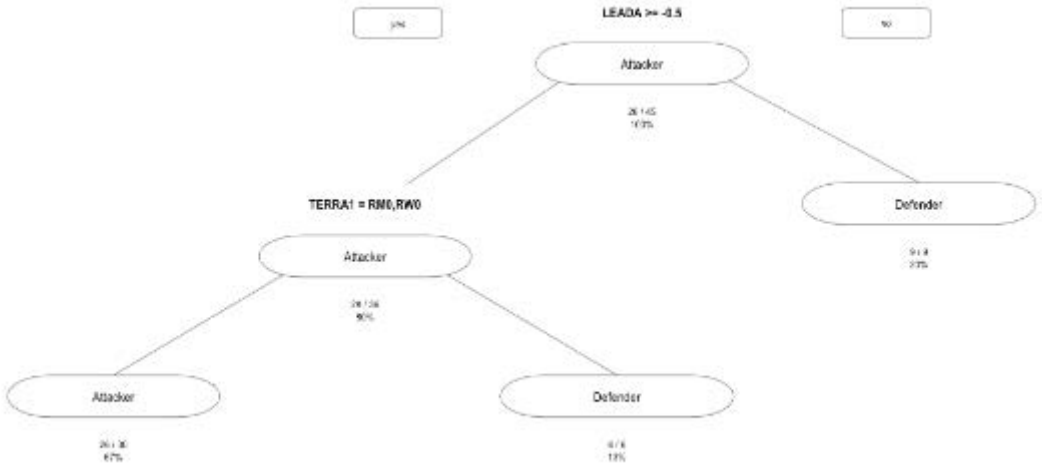
Figure 83. Tree Model 3.1: Battles between 1600 and 1697



2. Model 3.2 (1704–1799)

Model 3.2 consists of 63 battles from the eighteenth century. The misclassification rate for the training and the test dataset in this model is 0.13 and 0.22, respectively. In model 1.2 for the same subset, major splits were cavalry, artillery, force ratio, and leadership. In model 2.2 major splits were leadership and artillery ratio. However, in model 3.2, terrain emerged as an important variable with leadership.

Figure 84. Tree Model 3.2: Battles between 1704 and 1799



The primary splits for node one in this model are leadership, combat effectiveness, training, weather, and intelligence. The surrogate splits in this case for node one are combat effectiveness, weather, surprise, cavalry, and artillery ratio.

3. Model 3.3 (1800–1893)

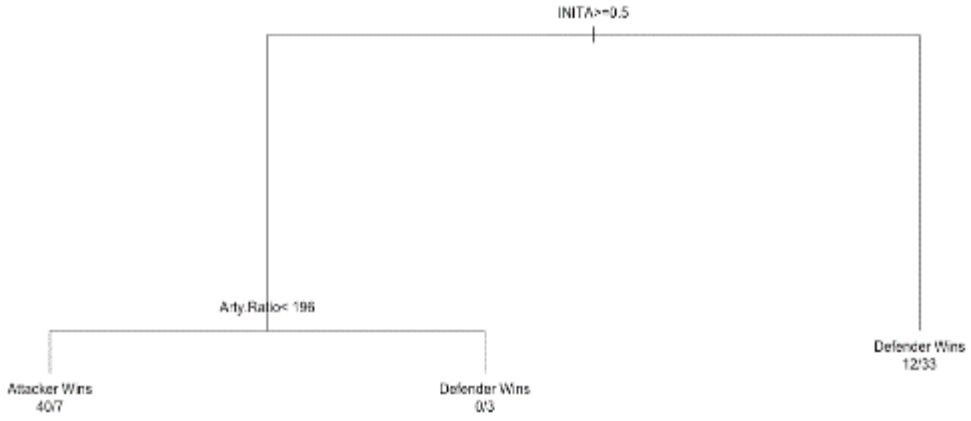
Model 3.3 consists of nineteenth century battles from 1800–1893. Total battles in model 3.3 are 122. The training set includes 76 and the test set includes 46 battles. The misclassification rate for this model is exactly same as model 2.3; for the training set it is 0.105 and the test set it is 0.369. The splits in this case are also the same: leadership and cavalry.

Figure 85. Tree Model 3.3: Battles between 1800 and 1893

4. Model 3.4 (1900–1920)

Model 3.4 consists of battles between 1900 and 1920. There are a total of 125 battles in the dataset. The misclassification rate for the training and the test dataset is 0.2 and 0.03, respectively.

Figure 86. Tree Model 3.4: Battles between 1900 and 1920



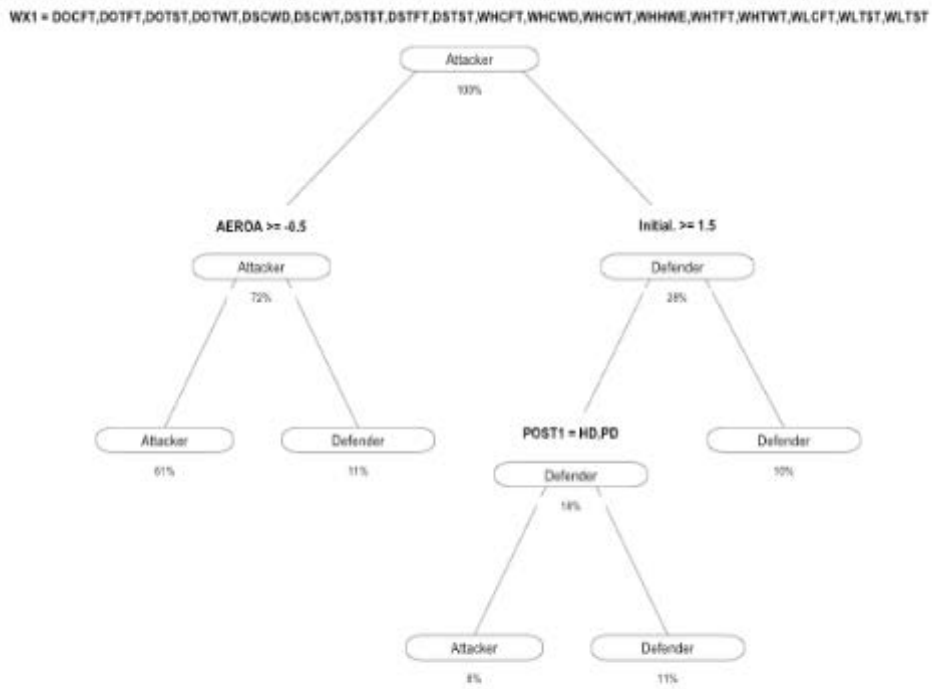
In model 2.4 on the same subset of data, splits were based on initiative and force ratio. In this case besides initiative, the model is splitting on artillery ratio. The primary

splits for node one in this model consist of initiative, leadership, weather, combat effectiveness, and terrain. The surrogate split for node one consists of weather, defensive posture, surprise, leadership and terrain.

5. Model 3.5 (1937–1945)

Model 3.5 consists of battles between 1937 and 1945. There are a total of 182 battles in this dataset. The misclassification rate for this model is 0.18 and for the training dataset is 0.33. The misclassification rate for the test dataset is higher than Models 1.5 and 2.5. The model is carrying out the first split on weather and second splits on air superiority and initial force ratio.

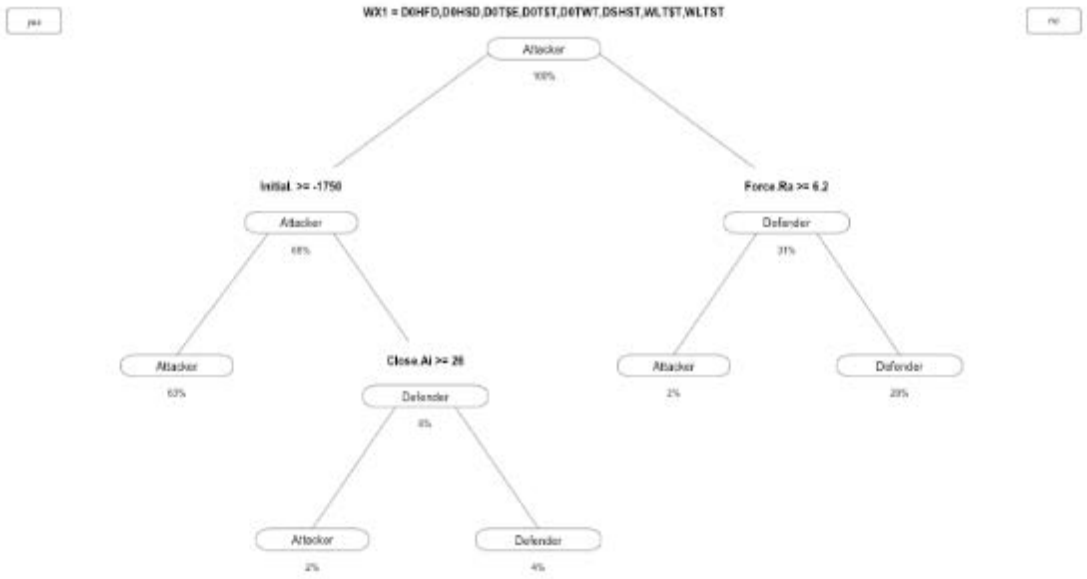
Figure 87. Tree Model 3.5: Battles between 1937 and 1945



6. Model 3.6 (1951–1982)

Model 3.6 consists of battles between 1951 and 1982. The variation in dataset results in the highest misclassification for the test dataset. From 1951 to 1982, there is a tremendous change in technology and warfare. The misclassification rate for the training dataset is 0.019 and the test data is 0.84. Model 2.6 was taking into account only air superiority, whereas in this model splits are based on weather, close air support, initial, and force ratio.

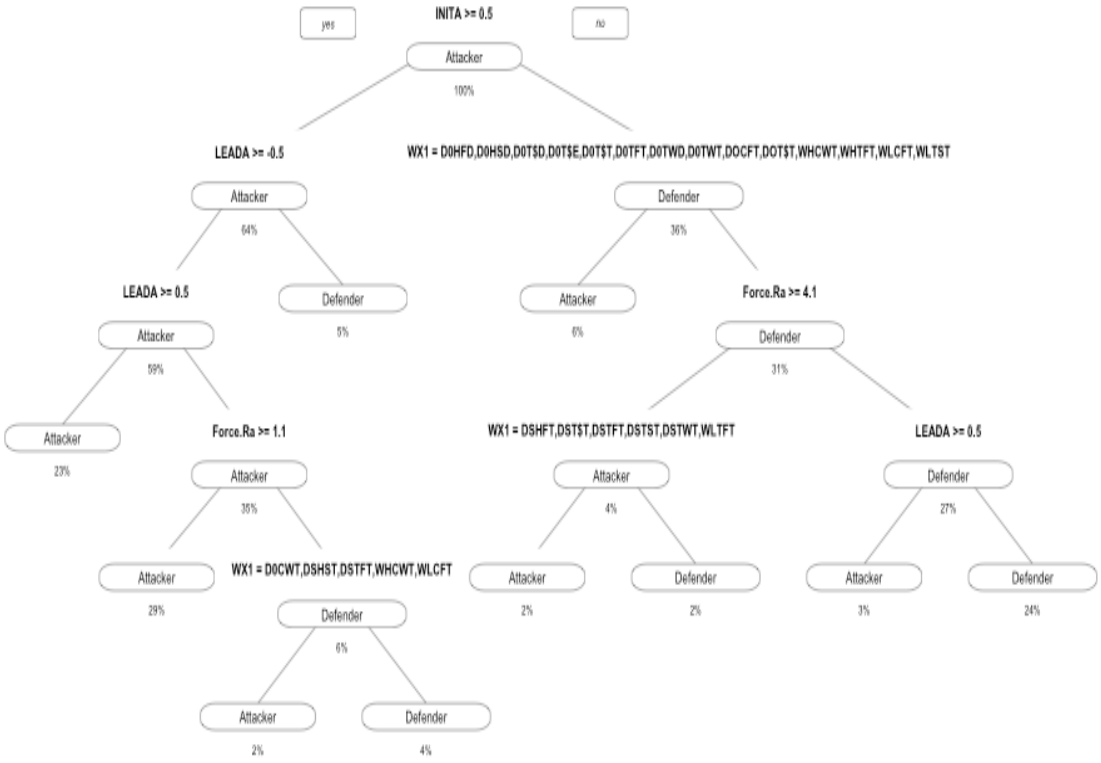
Figure 88. Tree Model 3.6: Battles between 1951 and 1982



7. Model 3.7 (1600–1982)

Model 3.7 consists of a complete dataset between 1600 and 1982. In this dataset there are 618 battles. The purpose of this model is to determine important variables. The misclassification rate for this model is 0.13. The misclassification rate for this model is lower than the misclassification rate of Models 1.7 and 2.7.

Figure 89. Tree Model 3.7: Battles between 1600 and 1982



E. ANALYSIS OF THE MODELS

Mathematical modeling for the outcome of the battles based on different types of predictor variables was carried out in the previous paragraphs. The last two models (Model 2 and 3) are systematically built on the first model by adding additional predictor variables.

The simplest form of the model is Model 1 that only considers objective variables such as force, tank, artillery, initial force, cavalry, and close air support ratios. Model 2 is an extension of Model 1 in which additional relative and subjective variables such as momentum, air superiority, initiative, surprise, intelligence advantage, leadership, and combat effectiveness have been added. Model 3 consists of the objective and the relative variables included in Model 2 along with three additional relative variables, which are logistics, training, and technology advantages. Model 3, besides these additional variables, also considers weather and terrain conditions.

All three models yield different misclassification rates for the test and the training datasets for different time periods. In terms of misclassification rates for the subset consisting of all the battles between 1600 and 1982, Model 3 is best, followed by Model 2, and Model 1. The misclassification rate decreases by adding relative variables to Model l. The misclassification rate for both the test and the training dataset for each type of model is given in Table 64.

Table 64. Misclassification Table

MISCLASSIFICATION TABLE							
Period	Subset	Model1		Model 2		Model 3	
		Train	Test	Train	Test	Train	Test
1600–1697	1	0.25	0.20	0.18	0.10	0.18	0.10
1704–1799	2	0.24	0.33	0.15	0.14	0.13	0.22
1800–1899	3	0.18	0.48	0.10	0.36	0.10	0.36
1900–1920	4	0.24	0.19	0.21	0.00	0.20	0.03
1937–1945	5	0.18	0.18	0.23	0.27	0.18	0.33
1951–1982	6	0.13	0.66	0.15	0.38	0.01	0.84
1600–1982	7	0.23		0.16		0.13	

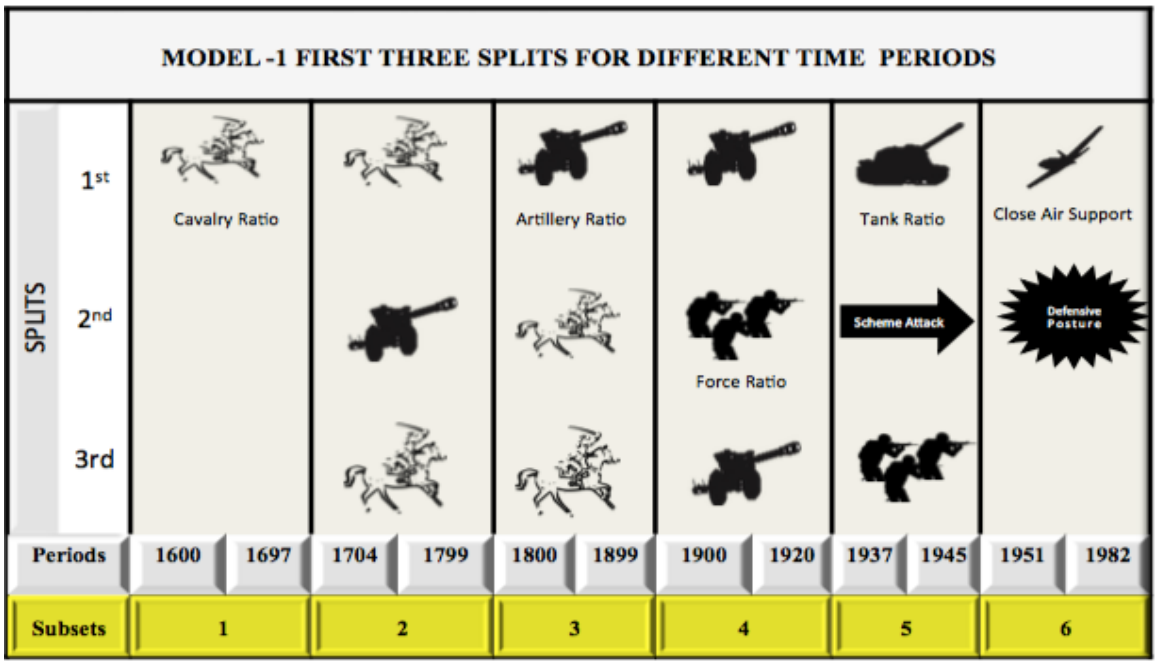
Based on the misclassification rates, Model 2 appears to be more accurate and resilient across different time periods. Model 1 has a higher misclassification rate than the other two models. Model 3 has a lower misclassification rate for the first four subsets. However, the misclassification rate for subset five and six in the case of Model 3 is higher than Model 2. Besides mathematical modeling and statistical analysis, it is pertinent to identify important variables that have been instrumental in the outcome of battles. In order to identify these important variables, a separate analysis of all three models has been carried out. The first three splits for each subset have been considered in order to ascertain the most important factors for winning battles.

1. Analysis Model 1 (Objective Variables)

Model 1 accounts for objective variables. In this model, there are six subsets that represent different time frames between 1600 and 1982. The variables selected in Figure 90 are based on the first three splits. In the case of subset one there is only one split based on cavalry. The split also corresponds to many qualitative analyses of warfare in which several historians identified cavalry as an important factor in winning battles. Another important aspect, which can be identified from the splits and from Figure 90, is that warfare has been evolving over time.

In the seventeenth century, cavalry was the prominent force for deciding the outcome of battle. In the eighteenth century with the development of modern artillery pieces, artillery gained prominence along with cavalry. In the nineteenth century, artillery changed the trend and became significant for deciding the outcome of battles. The domination of artillery continued between 1800 and 1920. However, between 1900 and 1920, force ratio also started to play a significant role in the outcome of battles. This was the era of trench warfare, where both artillery and force ratio were required to achieve victory. Between 1937 and 1945 (overlapping with World War II), tanks gained prominence along with scheme of maneuver and force ratio. During the period between 1951 and 1982, air superiority in theater of operation gained significant importance. This era includes battles between Arabs and Israel. With development in air warfare, the emphasis on the defensive scheme also gained prominence due to inclusion of the third dimension.

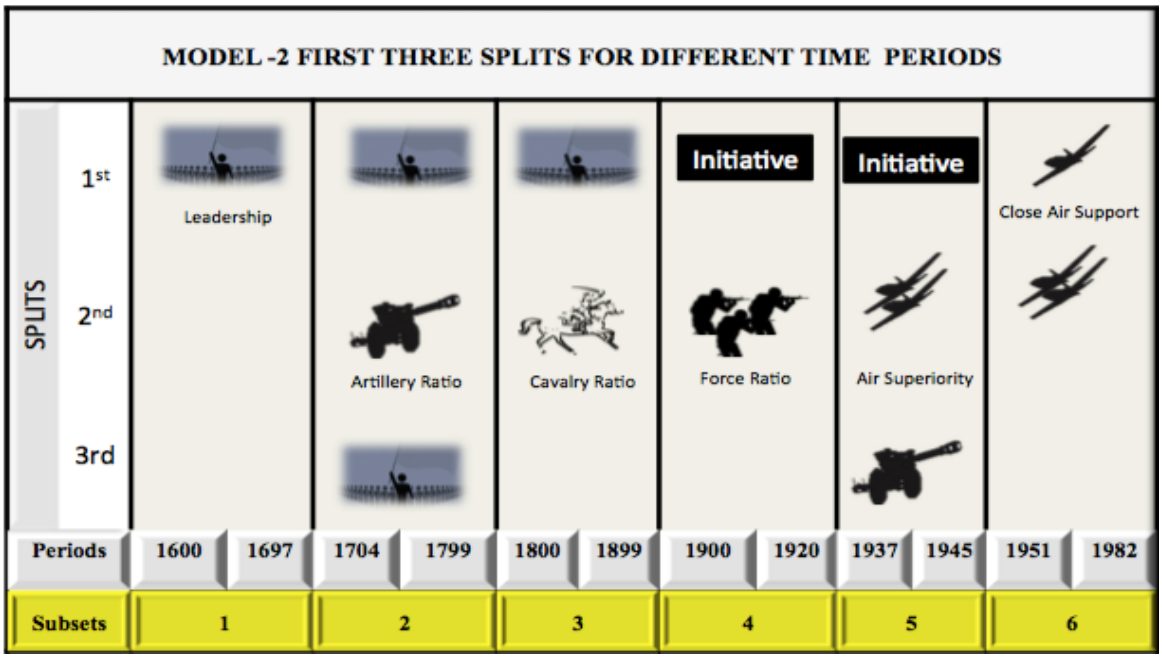
Figure 90. Model 1: First Three Splits



2. Analysis Model 2 (Objective and Relative Variables)

Model 2, as discussed earlier, is built on Model 1 with additional relative variables such as momentum, air superiority, initiative, surprise, intelligence, leadership, and combat effectiveness. The duration for subsets in this case are the same as in Model 1. The graphical representation of the first three splits is given in Figure 91. With the integration of relative variables, leadership appears to be the most vital instrument in deciding the outcome of battles between the seventeenth and the nineteenth centuries. Artillery and cavalry also played their role during this era. However, from 1900 to 1945, initiative is ranked first in comparison to artillery and tank ratio, which were the most important variables in Model 1. Initiative can be correlated with tanks since tanks were widely used by attackers for gaining initiative. Air superiority between 1937 and 1945 emerged as an important variable and continued with its prominence until 1982. The development in science and technology has a direct impact on the important factors associated with warfare, which can be visualized in Figure 91. Figure 91 shows that with the development of modern aircraft, aircraft became one of the most important variables for winning the battles.

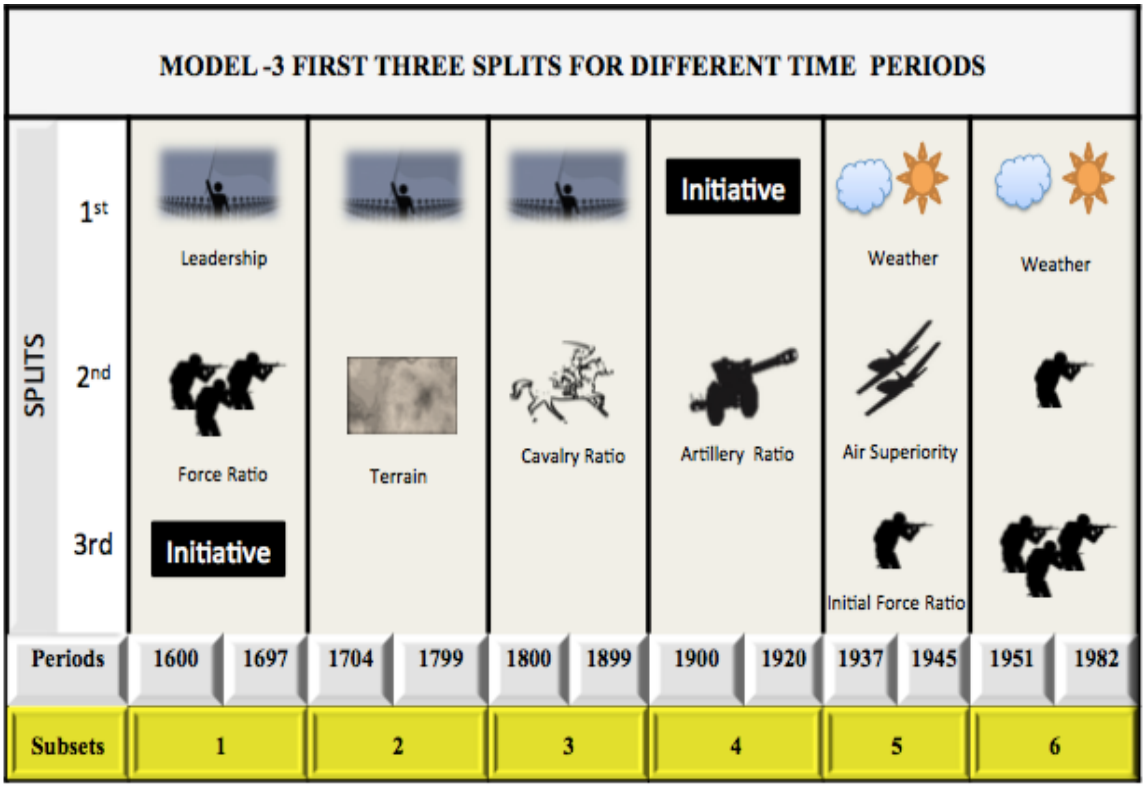
Figure 91. Model 2: First Three Splits



3. Analysis Model 3 (Objective Variables, Relative Variables, and Conditions)

Model 3 is built on Models 1 and 2, by incorporating three more relative variables along with terrain and weather conditions. Leadership remained important between 1600 and 1899. Another important factor during this era was cavalry. However, two other variables, which emerged during the seventeenth century, are force ratio and initiative. Terrain also emerged as an important variable during the eighteenth century. Between 1937 and 1982, weather and initial force ratio appeared as the important factors. The weather in those days had a relationship with air superiority and close air support. Clear weather was essential to gaining air support.

Figure 92. Model 3: First Three Splits



4. Important Variables

Warfare is transforming, and it is difficult to identify consistent winning factors. However, in each time frame there are different factors which can be identified based on the previous three models which can act as guidelines for military planners to plan future battles. Moreover, the battles of the recent past have more relevance to future battles as compared to past battles.

Table 65. Important Variables

IMPORTANT VARIABLES IN WINNING BATTLES	
Period	Important Variables
1600–1697	Leadership, Cavalry, Force Ratio, and Initiative
1704–1799	Leadership, Cavalry, Artillery, and Terrain
1800–1899	Cavalry, Leadership, and Artillery
1900–1920	Artillery, Force Ratio, and Initiative
1937–1945	Air Superiority, Tank Ratio, Force Ratio, Initial Force Ratio, Artillery, Scheme of Attack, and Weather
1951–1982	Close Air Support, Air Superiority, Force Ratio, Initial Force Ratio, Defensive Posture, and Weather

The important variables previously mentioned are selected based on the battles given in the dataset. These variables do not guarantee winning battles. Rather, they provide guidelines for military campaign planners for organizing their campaigns based on mathematical analysis of historical battles. In addition, these variables also validate the principles of war discussed in earlier chapters.

F. MODELLING FOR IRREGULAR WARFARE

Irregular warfare, as discussed earlier, is different in nature from conventional warfare. Moreover, quantified and well-documented information on various battles is less available. All three datasets analyzed during the course of this study contain information in different templates and are mainly focused on the social side of warfare instead of the military aspects. However, this endeavor is carried out to identify indicators of irregular conflicts. For identifying potential indicators of irregular conflicts, information from different datasets is combined and analyzed.

During the process, the number of events occurring in each country between 2006 and 2015 from the Armed Conflict Location and Event Data Project (ACLED) dataset is

taken as a response variable.[151] Predictor variables initially consist of 20 different World Bank development[152] and governance indicators[153] that were later reduced to five. These are Gross Domestic Product (GDP) (growth), GDP (current), Government Effectiveness, Political Stability, and Employment to Population Ratio. Moreover an additional predictor variable, status on the Fragile State Index managed by Fund For Peace,[154] was also incorporated. The Fragile State Index itself is composed of 12 different variables. These variables are demographic pressure, refugees, and internally displaced persons (IDPs), group grievances, human flight, uneven development, poverty and economic decline, legitimacy of the state, public service, human rights, security apparatus, factionalized elites, and external intervention.[155]

It is pertinent to discuss the scale of these predictors variable for the purpose of clarity to readers. Government Effectiveness is on ascending scale the higher the more number more effective is government. For example, Government Effectiveness of South Africa is 69.02 and Sudan is 11.21, which means the government of South Africa is more effective than Sudan. Political Stability is also on an ascending scale like Government Effectiveness. In the case of South Africa, the Political Stability score is 47.11, while in Sudan it is1.93, which means that South Africa is more politically stable than Sudan. Fragile state index is on a descending scale: the smaller the number, the better it is. The countries that are highly stable, such as Norway, have a very low number. The Fragile State index for 2015 has a score of 20.8 for Norway.

The purpose of our Irregular Model is to determine indicators that associated with political violence/irregular warfare using statistical techniques.

---

151 Clionadh Raleigh, Linke Andrew, Hegre Havard and Karlsen Joakim, "Introducing ACLED: An Armed Conflict Location and Event Dataset Special Data Feature," *Journal of Peace Research* 47, no. 5 (2010): 651–660.

152 World Bank Group, ed. *World Development Indicators 2012*. Washington, D.C.: World Bank Publications, 2012.

153 Daniel Kaufmann, Kraay Aart, and Massimo Mastruzzi, "The Worldwide Governance Indicators: Methodology and Analytical Issues," *Hague Journal on the Rule of Law* 3, no. 02 (2011): 220–246.

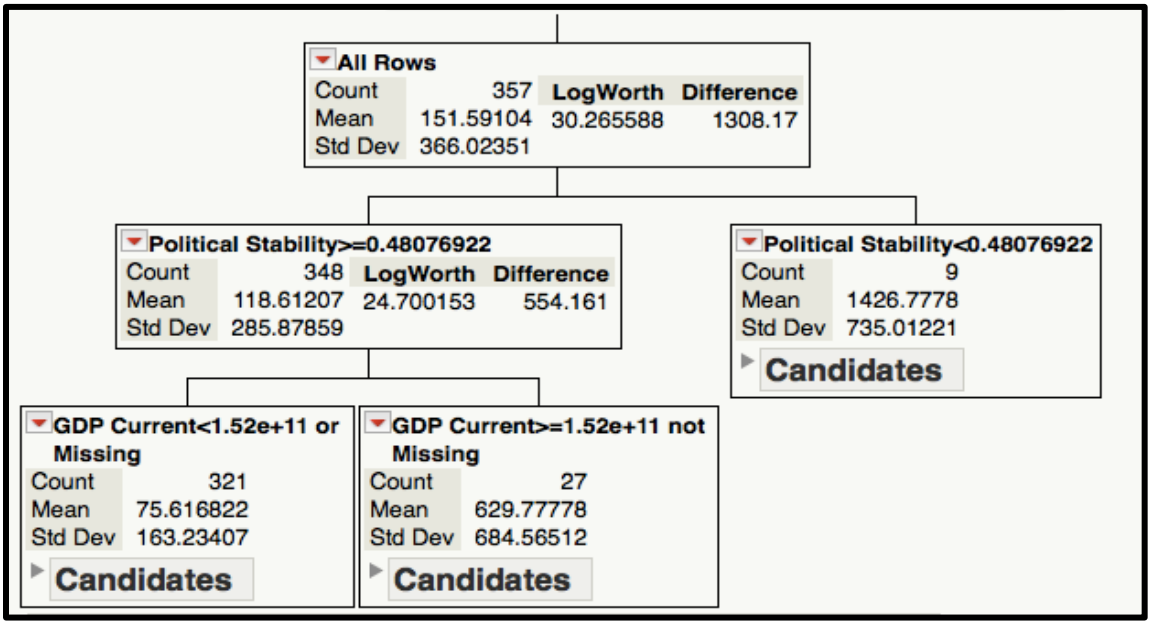
154 Foreign Policy, "Fund for Peace.(2006)," *Failed States Index* .

155 Ibid.

1. Irregular Warfare Model 1

The Irregular Warfare Model (IW Model 1) is built using a decision tree in JMP Pro by SAS.[156] The model uses the number of events of violence as the response variable. Predictor variables in this model include Fragile State Index status, Gross Domestic Product (current), Gross Domestic Product (growth), Government Effectiveness, Employment to Population Ratio, and Political Stability. The R Squared for the training dataset is 0.475 and for the validation dataset is 0.495. The tree splits based on two variables, political stability and GDP (current).

Figure 93. Model-1: Irregular Warfare

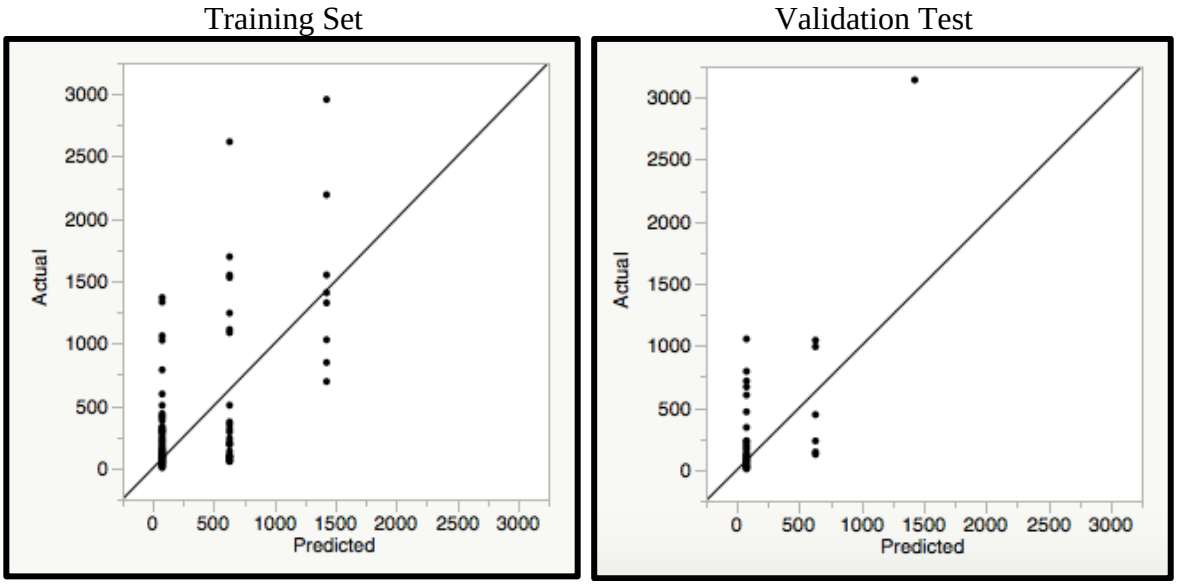


The Root Mean Square Error (RMSE) for the training dataset is 264.80 and for the validation dataset is 262.07. The JMP software enabled us to find the contribution of various predictor variables in the construct of a model. In this case, there are only two variables based on which this model is built and that are Political Stability with 0.66 percentage of contribution to model and GDP (current) with 0.33 percentage of contribution to model. IW Model 1 suggest that by increasing Political Stability and

[156] Pro, *v10*

improving GDP, the number of violent events may be reduced. Irregular warfare is dynamic and varies in different situations. As a result, it is difficult to capture all underlying causes of this kind of conflict in a small model. IW Warfare Model 1 focuses on very limited variables and indicators due to the scope of this research and limited availability of time.

Figure 94. Actual Versus Predicted Plot Model-1 for Irregular Warfare



2. Irregular Warfare Model 2

Irregular Warfare Model 2 (IW Model 2) is developed using a boosted tree on the same dataset. Boosting is the method of building a large, additive decision tree by fitting a sequence of smaller trees. Each of the smaller trees is fit on the scaled residuals of the previous tree. The trees are combined to form the larger final tree. The process can use validation to assess how many stages to fit, not to exceed the specified number of stages. The tree at each stage is short, typically one to five splits.[157] After the initial tree, each stage fits the residuals from the previous stage. The process continues until the specified number of stages is reached, or, if validation is used, until fitting an additional stage no

[157] A. JMP and Marcel Proust, "Specialized Models," 2013.

longer improves the validation statistic. The final prediction is the sum of the estimates for each terminal node over all the stages.[158]

Model 2 has higher R Squared for both the training and the validation sets. Moreover, Root Mean Square Error (RMSE) is also lower then IW Model 1. The R Squared for the training set in IW Model 2 is 0.851 and for the test set is 0.801. The Root Mean Squared Error for the training set is 146.23 and the validation set is 135.0652. The number of layers in the boosted tree model is 115, splits per tree are three, and the learning rate is 0.1. Learning rate is a number such that $0 < r \leq 1$.[159] Learning rates close to one result in faster convergence on a final tree, but also have a higher tendency to over fit data.[160] According to Friedman, empirical studies have shown that a learning rate of .1 or less usually leads to better models (with better predictive validity).[161] The learning curve is continuously rising, and at tree seven it crosses the R Squared achieved by IW Model 1.

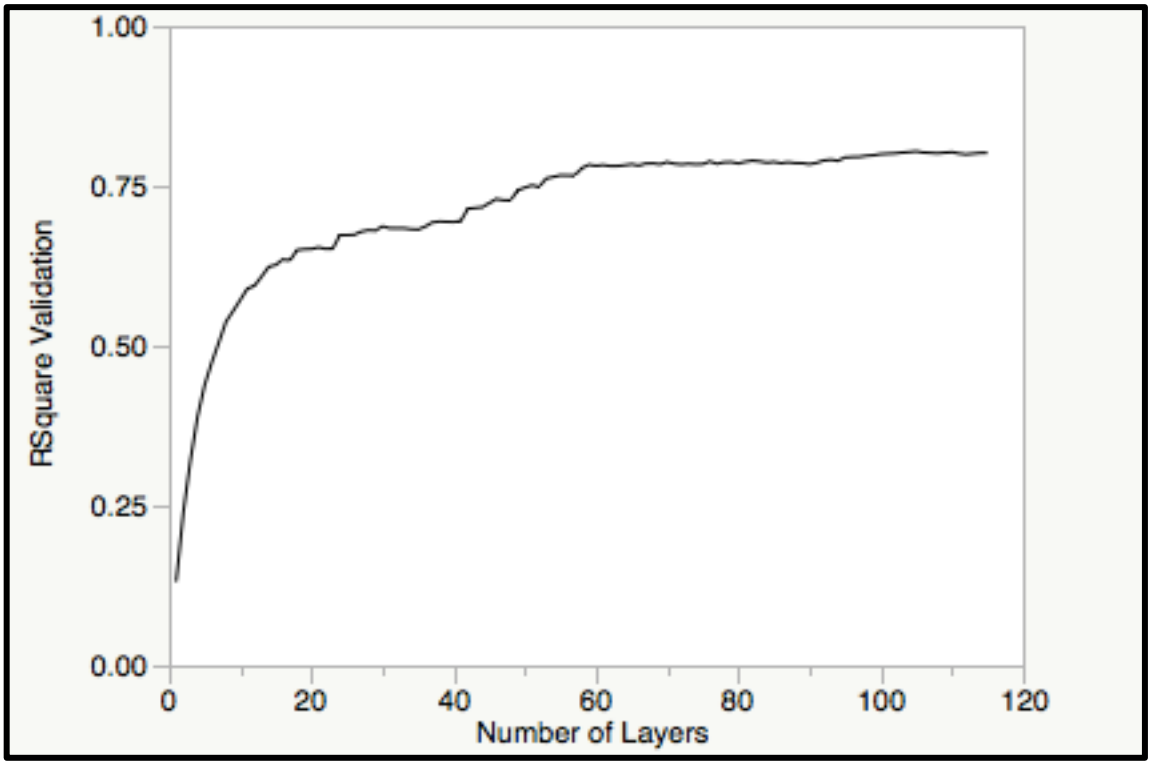
158 Ibid.

159 Ibid.

160 Ibid.

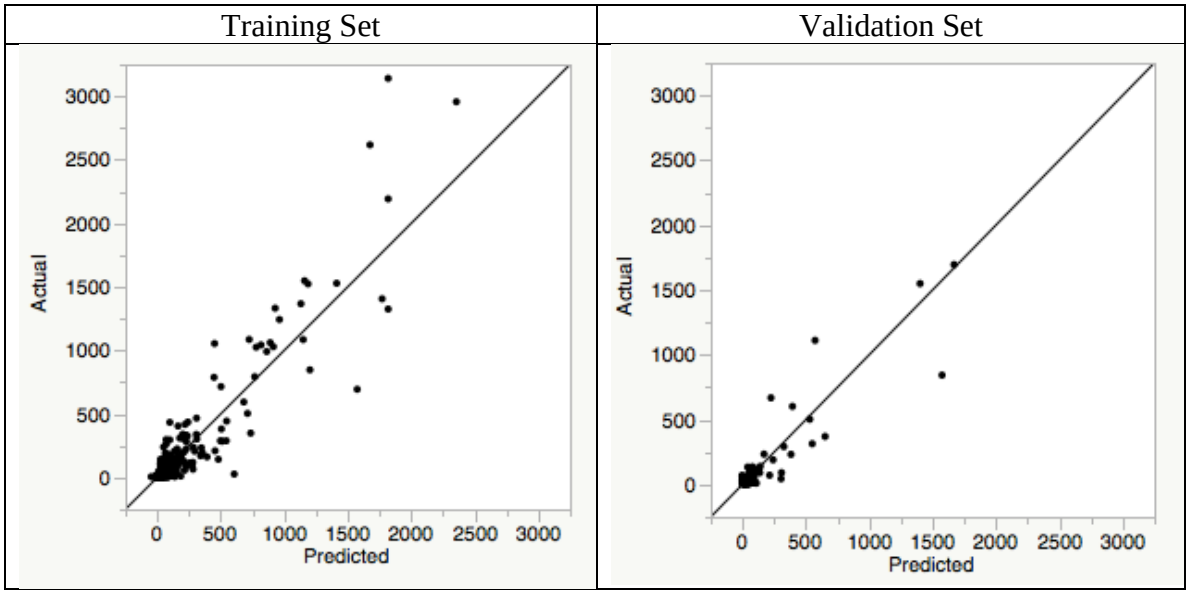
161 Jerome Friedman, Hastie Trevor, and Tibshirani Robert, "Additive Logistic Regression: A Statistical View of Boosting," *The Annals of Statistics* 28, no. 2 (2000): 337–407.

Figure 95. Cumulative Validation Model-2: Irregular Warfare



Besides Political Stability and GDP (current) other variables contribute to the model. With respect to the contribution of various variables, IW Model 2 includes 0.6 proportion contribution of Political Stability, 0.2 percentage contribution of GDP (current), 0.05 of Employment to Population Ratio, 0.05 of Fragile State Index, 0.04 of GDP (growth), and 0.07 percentage of Government Effectiveness. Although this model utilizes more variables for predicting the outcome, it is still unable to capture all of the details required for comprehending irregular warfare.

Figure 96. Actual Versus Predicted Plot Model-2 for Irregular Warfare



VII. FINDINGS AND FUTURE STUDIES

Chapter VII focuses on important findings of this research. Moreover, it lists future possible studies that can be carried out on the subject.

A. IMPORTANT FINDINGS

Warfare is a complex phenomenon. It is difficult to sum up details of warfare in a solitary study. The variables involved in warfare are difficult to measure, since it involves both qualitative and quantitative measurements. Intangible factors such as leadership, momentum, and intelligence are not only difficult to measure, but their scale may also differ based on the method of measurement. Moreover, as it is commonly said, the victor writes history. This impacts the collection of data related to warfare. Data collected after a culmination of war at times tend to be biased in favor of the victor. These biases affect subsequent analysis, and sometimes the tactical brilliance of soldiers is shadowed by strategic failures of their leaders.

The data collection in irregular warfare tends to be more difficult than for conventional warfare. The numbers of participants in irregular warfare can be greater than conventional warfare, and there is no archiving of data by non-state actors. However, despite all of these drawbacks, this research has endeavored to use authenticated, well-documented, and credible datasets to reduce the effect of biases and get insight into warfare. Moreover, well-established principles of wars have been intrinsically correlated with various variables to find mathematical solutions to the complex process of winning battles. This research analyzed four different datasets containing information on different kinds of battles.

B. IMPORTANT FINDINGS ON CONVENTIONAL WARFARE

The CDB90G dataset was compiled by the Center for Army Analysis. This dataset includes 664 battles between the seventeenth and twentieth centuries. The first battle in the dataset is the battle of Nieuport, which was carried out during the Netherlands war for independence at Spanish Flanders. The last battle in the dataset is the

1982 battle of Lebanon, which took place at Bekka. This dataset focused on conventional battles. The major findings related to conventional battles from the CDB90G dataset are the following:

1. In military planners' opinions, offense is the best form of defense. The operational research has converted this opinion to fact by mathematically establishing that the attacker tends to have an advantage over the defender. The analysis reveals that the attacker is successful 60.93% of the time, the defender 32.6%, and 6.62% of the battles result in a draw. This important fact reveals that while planning for future battles, emphasis should be laid on offensive strategy as compared to defensive.

2. Climatic conditions played a major role in deciding a time frame of battles. From the analysis of the dataset, it is revealed that military leaders prefer to carry out campaigns in moderate temperatures. In the countries in which weather is hot, most campaigns were carried out in winter. In the countries with cold temperatures, campaigns were mostly carried out in summer. We can confidently establish that climate based on geographical location affects campaign planning.

3. Force ratio has always been of paramount importance to military leaders. The rule of thumb 3:1 attacker-to-defender force ratio is applicable worldwide. The question under discussion is whether this established rule has some empirical evidence or if it is hypothetically drawn. From analysis of the dataset, we conclude that there are only 121 battles in which attacker-to-defender ratio was equal to or more than 3:1. Of these 121 battles, the attacker won 84 battles and the defender won 37 battles. The percentage of attacker victories when force ratio is equal to and more than 3:1 is 69.4%. However, another fact that can be established from analysis of force ratio is that most of the battles (317 in this dataset) were fought below a 3:1 ratio. This indicates that mustering a 3:1 force ratio is a difficult task in wars since the commander will always be short of manpower. This limitation of manpower can be overcome by imparting rigorous training to forces.

4. From the analysis of the CDB90G dataset, it is revealed that the presence of tanks on the battlefield has a direct impact on the outcome of wars. There are 300 battles in which tanks were used. Of these 300 battles, the attacker won 199 battles when both sides had tanks and won 56 when the defender did not have tanks. This confirms that tanks play a vital role, especially in offensive battles. It supports the adage that "heavy forces beat light forces."

5. Artillery is an important component of all armies. From the analysis of the dataset, it is revealed that if artillery ratio is in favor of the attacker, the chances of winning the battle increases.

6. The analysis reveals that warfare remains dynamic throughout history, and technology has directly impacted warfare. During the fifteenth century, leadership, cavalry, force ratio, and initiative dominated warfare. In the sixteenth century, leadership, cavalry, artillery, and terrain became significant variables for winning battles. In the nineteenth century, cavalry, leadership, and artillery remained as the most significant variables for deciding the outcome of battles. In the twentieth century, due to rapid development in science and technology, different variables dominated warfare. We can segment the twentieth century into three different portions. The first period, between 1900 and 1920, is dominated by artillery, force ratio, and initiative. The second period, between 1937 and 1945, is dominated by air superiority, tank ratio, force ratio, initial force ratio, artillery, scheme of attack, and weather. The third period, between 1951 and 1982, is dominated by close air support, air superiority, force ratio, initial force ratio, defensive posture, and weather.

7. Weather does not appear in earlier battles as an important factor, but for later periods of the twentieth century it attained significant importance. The reason for this is the use of air power in later periods; airplanes can be hindered by weather conditions, so this variable gained importance.

8. The dataset contains battles until 1982. After 1982, there is major development in technology; the dataset does not contain that data. It would be interesting to look at the battles and conflicts after 1982 to see how satellites, drones, and other weapons impact modern battles.

9. The most import conclusion that we can draw from the analysis of this dataset is that warfare has been changing with time, and in order to win battles commanders must keep their forces equipped with the latest technologies. Moreover, human factors such as leadership and initiative will keep playing a role in future battles. There is no substitute for human decision-making so far.

C. IMPORTANT FINDINGS ON IRREGULAR WARFARE

Irregular warfare is a complex phenomenon since it involves a wide variety of different groups and reasons for conflict, and varies based on geographical location and type of participants. Historically, irregular warfare has remained one of the major forms of warfare, yet there is very little mathematical analysis of it. In the recent past, irregular warfare has dominated the globe because of ongoing conflicts in Iraq and Afghanistan. Most military analysts think that future conflicts will be dominated by this kind of

warfare.[162] In order to understand irregular warfare, we analyzed three different datasets to reveal trends in irregular warfare and determine related factors for these kinds of conflicts. Important findings based on the analysis of these datasets are discussed in the succeeding paragraphs.

1. Findings Based on Analysis of ACLED

a. ACLED is a collection of data on political violence focused on Asian and African countries. Analysis of the dataset reveals that most events (36,809) that took place were *battle no change of territory*. The second greatest number of events (33,065) that occurred were *violence against civilians*. The *battle no change of territory* is an activity which seems to be common during protracted irregular conflicts. These results also validate an important cliché that a population is the center of irregular warfare. In all of these conflicts, civilians were the major targets of all the actors. This reflects how important winning or suppressing the support of the population is in irregular warfare.

b. Another trend that can be observed from ACLED is that in irregular warfare, conflict tends to occur in close vicinity to international borders. We can infer that insurgents require support and safe havens from across the border. This also suggests that if borders are not safeguarded properly, insurgents can use them to their own advantage. This trend is obvious in Afghanistan, Cambodia, Haiti, Laos, Myanmar, Nepal, Pakistan, and in African countries.

c. Seasonality effects are very prominent in irregular conflicts. Most of the time insurgents carry out operations during a time of year that is moderate. This seasonality effect is visibly identified in Afghanistan, Haiti, Laos, and Cambodia. In Afghanistan, the campaign season is mostly in summer since winters in Afghanistan are harsh and do not support campaigns.

d. Irregular warfare is heterogeneous as far as the actors involved are concerned. The complexity of irregular warfare has multiplied due to the presence of different groups. It has been established through analysis of data that there are normally more than 10 actors involved in irregular conflicts. In cases such as Afghanistan and Pakistan, the situation is more complex due to the involvement of the large number of players.

e. In all the countries that were analyzed, except Nepal, there are foreign participants involved in the conflicts. The participants are

[162] John Arquilla and Ronfeldt David, *Networks and Netwars: The Future of Terror, Crime, and Militancy* (Santa Monica, CA: Rand Corporation, 2001).

on the sides of both the government and the insurgents. For example, in Afghanistan there is the presence of al-qaeda with the Taliban and International Security Assistance Force (ISAF) with the Afghan national government.

f. From the ACLED dataset, we observed the trend that events are not confined to geographical boundaries of the country in which conflict or political violence is occurring. Sometimes, events occur in neighboring countries. This tendency of events to occur in neighboring countries was observed in all countries except Nepal.

g. In all the countries that were analyzed in ACLED, insurgent movements tend to follow a definite pattern. In this pattern, initially they start with fewer events, then reach a peak, and then subsequently die down.

h. There is a trend of violence in countries with low income, fragile government, low GDP, and poor economic conditions. These factors favor the instigating and flourishing of violence.

i. Africa has the highest number of events with an exponential growth of violence. This trend of violence in Africa is growing. This may be due to the growing population of Africa and inability of the governments to incorporate the growing population in the society. From the military perspective, these circumstances provide valid reasons to explore Africa as a potential future conflict area for irregular warfare. The data also showed a presence of Islamic militants in Africa. These militants have a presence in different African countries, and they can use Africa as the new base to initiate terrorism across the globe. In the dataset, Africa was observed in the holistic approach as a continent. If each country is analyzed separately, more insight could be obtained.

j. The ACLED dataset was modeled using different World Bank indicators to determine factors that are associated with irregular conflicts. During modeling, it was established that for predicting irregular conflicts governance indicators play a more significant role in comparison to development indicators. Important factors such as political stability, GDP, employment to population ratio, and government effectiveness can reduce the possibility of these kinds of conflicts. In other words, if efforts are made to improve these factors, we may be able to considerably reduce violence.

k. Research identified Africa as a major trouble spot, and the trend seems likely to continue into the future as well. In order to avoid future irregular conflicts, the world community along with international aid agencies should focus on the previously-mentioned variables to minimize future conflicts in Africa.

2. Findings Based on COW

a. From COW it is established that most numbers (250) of intra-state war are carried out for central control. The second most number of wars were over local issues (161). The maximum number of conflicts occurred in Africa. The highest numbers of conflicts (73) in Africa are also for central control. The two datasets, ACLED and COW, are verifying the trend towards political instability in Africa.

b. The number of casualties in civil war for central control and civil war for local issues are within the same range and are more than regional internal war and inter-communal wars.

c. The duration of civil war for central control and civil war for local issues is more than regional internal and communal wars. The mean length for civil war for central control is 769.06 days, and for civil war over local issues it is 870.94 days. Whereas, the mean for regional internal war is 225.54 and inter-communal war is 379.05. This reflects that civil war for central control and over local issues tends to be longer in duration.

d. The casualties and duration of war established that most of the casualties happened within the first 1,000 days, and after that, the number of casualties did not increase substantially for all kinds of war. We can conclude that during most wars, both sides initially suffer more casualties, and then subsequently the number of casualties tends to decrease. Another inference that can be drawn based on this mathematical analysis is that the initial days of war are more intense, and then subsequently activities tend to be less significant in terms of casualties.

e. The analysis of non-state wars of the COW project indicates that the maximum number of non-state wars occurred in Africa and Asia. This confirms our finding based on ACLED that Africa has the potential to be a breeding space for terrorists and insurgents due to its political instability, poor government, increasing population, and potential of young manpower to be exploited by extremists.

f. The casualties in non-state wars are higher than in other types of wars. The mean for casualties in non-state war is 8,269.6, which is higher than other types of conflict in the COW dataset. This indicates that non-state war tends to be casualty intensive.

g. The maximum number of casualties in non-state war are typically within the first 300 days; after that there are fewer casualties.

3. Finding Based on Uppsala

a. The Uppsala non-state conflict dataset indicates that most events (520) are carried out by organization level three, which are informally organized groups. In other words, these are groups that

operate in loosely connected networks and have the ability to connect for carrying out many kinds of activity.

b. The duration of conflict for organization level three, which are loosely connected or informally organized groups, is more than for the other two organizations. The mean duration for organization level three is 975.17 days. This means that the loosely organized group can be involved in long conflicts while keeping itself hidden from law enforcement agencies. From a military standpoint, it suggests that in order to tackle these groups, a strategy should be formulated to observe their network and strike at critical nodes that incapacitate their ability to become active.

c. The dataset also showed that organizations of all three levels that are discussed in Chapter V operate on a system of networks. These networks are not essentially confined to a particular geographical area; rather, they tend to spread out in different areas. The one way to effectively counter these networks is to identify the main hubs on which other nodes are dependent for operation.

d. The analysis of Uppsala one-sided violence indicates that most (389) of the one-sided violent conflicts happened in Africa. This trend has been identified in all three different datasets that we examined.

D. FUTURE STUDIES

As already stated, warfare is an extremely complex phenomenon. It is difficult to capture all of the details in one research paper. This research is a small endeavor to gain insight into winning aspects and trends of warfare based on available data. There are many other possible opportunities for future research that will further enhance our understanding of warfare. In subsequent paragraphs, we discuss these opportunities with the prospect that application of operation research techniques will assist military planners to solve future conflicts.

1. Only a few of the variables in the CDB90G dataset have been explored in this research. There are possibilities of exploring more variables to identify what other factors can be added for predicting outcomes of battles. The research focuses on decision tree methodology for predicting outcome of battles; application of other techniques such as neural networks may give additional insights.
2. Another option for future research is simulating the data and then carrying out analysis of available data with simulated data to gain a more holistic picture of warfare.
3. The ACLED, COW, and Uppsala dataset are not explored completely. They have a lot of information that can be explored to get more insight

into irregular warfare. Specific study focused on each dataset will surely reveal interesting new information related to conflicts.

4. The COW and Uppsala datasets can be modeled to determine a mathematical relationship between casualties and the number of days of conflict. This study will provide military leaders with risk indicators associated with future campaigns.
5. Few development and governance indicator future were explored to determine relationships between violent activity, political instability, and socio-economic conditions. This is a vast area in which operations research can assist military leaders to more accurately predict future conflicts. This will give militaries an advantage to prepare and train for future conflict well in advance.

E. CONCLUSION

The research has endeavored to capture details of both conventional and irregular warfare from the perspective of a military and operations research analyst. However, the availability of data, time constraints, and the vastness of the subject remained the core issue. Thus, only selected variables could be explored. This research is just a first step on discovery on a most important road. This research can be taken as an initiation step that can be further explored using both quantitative and qualitative techniques to discover mysteries and myths related to warfare.

APPENDIX A.

The appendixes A to C consist of three different types of tree model based on the CDB90G dataset. These models are output results for readers who wanted to get complete statistical insight into models. Since Chapter VI only gives graphical representation of the tree models, these details are displayed here especially for the operational research students who are more interested in gaining statistical insight in comparison to ordinary military readers.

1. **MODEL 1.1**

1) root 27 9 Attacker Wins (0.6666667 0.3333333)
2) Cav.Ratio< 1.035714 19 4 Attacker Wins (0.7894737 0.2105263) *
3) Cav.Ratio>=1.035714 8 3 Defender Wins (0.3750000 0.6250000) *

2. **Model 1.2**

1) root 45 17 Attacker Wins (0.6222222 0.3777778)
2) Cav.Ratio>=0.05 38 12 Attacker Wins (0.6842105 0.3157895)
4) Arty.Ratio>=1.521237 10 1 Attacker Wins (0.9000000 0.1000000) *
5) Arty.Ratio< 1.521237 28 11 Attacker Wins (0.6071429 0.3928571)
10) Cav.Ratio< 0.8928571 7 1 Attacker Wins (0.8571429 0.1428571) *
11) Cav.Ratio>=0.8928571 21 10 Attacker Wins (0.5238095 0.4761905)
22) Force.Ratio>=0.827068 14 5 Attacker Wins (0.6428571 0.3571429) *
23) Force.Ratio< 0.827068 7 2 Defender Wins (0.2857143 0.7142857) *
3) Cav.Ratio< 0.05 7 2 Defender Wins (0.2857143 0.7142857) *

3. **Model 1.3**

1) root 83 40 Defender Wins (0.4819277 0.5180723)
2) Arty.Ratio< 0.7876244 18 3 Attacker Wins (0.8333333 0.1666667) *
3) Arty.Ratio>=0.7876244 65 25 Defender Wins (0.3846154 0.6153846)
6) Cav.Ratio< 0.3403922 8 1 Attacker Wins (0.8750000 0.1250000) *
7) Cav.Ratio>=0.3403922 57 18 Defender Wins (0.3157895 0.6842105)
14) Cav.Ratio>=1.104403 22 10 Attacker Wins (0.5454545 0.4545455)
28) Initial.Force.Ratio< 1.252513 11 2 Attacker Wins (0.8181818 0.1818182) *
29) Initial.Force.Ratio>=1.252513 11 3 Defender Wins (0.2727273 0.7272727) *
15) Cav.Ratio< 1.104403 35 6 Defender Wins (0.1714286 0.8285714) *

4. **Model 1.4**

1) root 95 43 Attacker Wins (0.5473684 0.4526316)

2) Arty.Ratio< 196 87 36 Attacker Wins (0.5862069 0.4137931)
4) Force.Ratio>=0.8366652 78 29 Attacker Wins (0.6282051 0.3717949)
8) Arty.Ratio< 0.9791667 18 3 Attacker Wins (0.8333333 0.1666667) *
9) Arty.Ratio>=0.9791667 60 26 Attacker Wins (0.5666667 0.4333333)
18) Arty.Ratio>=1.009225 41 12 Attacker Wins (0.7073171 0.2926829) *
19) Arty.Ratio< 1.009225 19 5 Defender Wins (0.2631579 0.7368421) *
5) Force.Ratio< 0.8366652 9 2 Defender Wins (0.2222222 0.7777778) *
3) Arty.Ratio>=196 8 1 Defender Wins (0.1250000 0.8750000) *

5. **Model 1.5**

1) root 133 51 Attacker Wins (0.6165414 0.3834586)
2) Tank.Ratio>=3.785833 48 7 Attacker Wins (0.8541667 0.1458333) *
3) Tank.Ratio< 3.785833 85 41 Defender Wins (0.4823529 0.5176471)
6) PRIA1=DE,RC 13 2 Attacker Wins (0.8461538 0.1538462) *
7) PRIA1=EE,FF 72 30 Defender Wins (0.4166667 0.5833333)
14) Force.Ratio>=1.389386 45 22 Attacker Wins (0.5111111 0.4888889)
28) Close.Air.Support.Ratio>=7.5 8 1 Attacker Wins (0.8750000 0.1250000) *
29) Close.Air.Support.Ratio< 7.5 37 16 Defender Wins (0.4324324 0.5675676)
58) Force.Ratio< 3.305437 29 14 Attacker Wins (0.5172414 0.4827586)
116) Arty.Ratio>=2.221951 7 1 Attacker Wins (0.8571429 0.1428571) *
117) Arty.Ratio< 2.221951 22 9 Defender Wins (0.4090909 0.5909091)
234) Close.Air.Support.Ratio< 0.06164384 9 3 Attacker Wins (0.6666667 0.3333333) *
235) Close.Air.Support.Ratio>=0.06164384 13 3 Defender Wins (0.2307692 0.7692308) *
59) Force.Ratio>=3.305437 8 1 Defender Wins (0.1250000 0.8750000) *
15) Force.Ratio< 1.389386 27 7 Defender Wins (0.2592593 0.7407407) *

6. **Model 1.6**

1) root 51 18 Attacker Wins (0.6470588 0.3529412)
2) Close.Air.Support.Ratio>=2.725877 25 1 Attacker Wins (0.9600000 0.0400000) *
3) Close.Air.Support.Ratio< 2.725877 26 9 Defender Wins (0.3461538 0.6538462)
6) POST1=DL,FD,PD 15 6 Attacker Wins (0.6000000 0.4000000) *
7) POST1=HD 11 0 Defender Wins (0.0000000 1.0000000) *

7. **Model 1.7**

1) root 618 217 Attacker Wins (0.6488673 0.3511327)
2) Tank.Ratio>=3.785833 111 13 Attacker Wins (0.8828829 0.1171171) *
3) Tank.Ratio< 3.785833 507 204 Attacker Wins (0.5976331 0.4023669)
6) PRIA1=0,DE,DO,EE,RC 74 13 Attacker Wins (0.8243243 0.1756757) *
7) PRIA1=FF 433 191 Attacker Wins (0.5588915 0.4411085)

14) Arty.Ratio< 0.3601619 72 21 Attacker Wins (0.7083333 0.2916667) *
15) Arty.Ratio>=0.3601619 361 170 Attacker Wins (0.5290859 0.4709141)
30) Force.Ratio>=1.329388 189 76 Attacker Wins (0.5978836 0.4021164)
60) Arty.Ratio< 156 182 69 Attacker Wins (0.6208791 0.3791209)
120) Close.Air.Support.Ratio>=13.8 16 2 Attacker Wins (0.8750000 0.1250000)
*
121) Close.Air.Support.Ratio< 13.8 166 67 Attacker Wins (0.5963855
0.4036145)
242) Close.Air.Support.Ratio< 0.06164384 128 45 Attacker Wins (0.6484375
0.3515625)*
243) Close.Air.Support.Ratio>=0.06164384 38 16 Defender Wins (0.4210526
0.5789474)
486) Close.Air.Support.Ratio>=0.8055556 25 11 Attacker Wins (0.5600000
0.4400000)
972) Close.Air.Support.Ratio< 2.1 13 3 Attacker Wins (0.7692308 0.2307692) *
973) Close.Air.Support.Ratio>=2.1 12 4 Defender Wins (0.3333333 0.6666667)
*
487) Close.Air.Support.Ratio< 0.8055556 13 2 Defender Wins (0.1538462
0.8461538) *
61) Arty.Ratio>=156 7 0 Defender Wins (0.0000000 1.0000000) *
31) Force.Ratio< 1.329388 172 78 Defender Wins (0.4534884 0.5465116)
62) Initial.Force.Ratio< 1.268086 159 78 Defender Wins (0.4905660 0.5094340)
124) Close.Air.Support.Ratio>=1.789583 18 5 Attacker Wins (0.7222222
0.2777778) *
125) Close.Air.Support.Ratio< 1.789583 141 65 Defender Wins (0.4609929
0.5390071)
250) Tank.Ratio< 0.8546171 132 64 Defender Wins (0.4848485 0.5151515)
500) Force.Ratio>=0.8026549 87 40 Attacker Wins (0.5402299 0.4597701)
1000) Force.Ratio< 0.8974898 18 4 Attacker Wins (0.7777778 0.2222222) *
1001) Force.Ratio>=0.8974898 69 33 Defender Wins (0.4782609 0.5217391)
2002) Arty.Ratio< 0.983871 15 5 Attacker Wins (0.6666667 0.3333333) *
2003) Arty.Ratio>=0.983871 54 23 Defender Wins (0.4259259 0.5740741)
4006) Force.Ratio>=1.128129 23 9 Attacker Wins (0.6086957 0.3913043)
8012) Force.Ratio< 1.239983 14 3 Attacker Wins (0.7857143 0.2142857) *
8013) Force.Ratio>=1.239983 9 3 Defender Wins (0.3333333 0.6666667) *
4007) Force.Ratio< 1.128129 31 9 Defender Wins (0.2903226 0.7096774) *
501) Force.Ratio< 0.8026549 45 17 Defender Wins (0.3777778 0.6222222)
1002) Force.Ratio< 0.4477727 7 2 Attacker Wins (0.7142857 0.2857143) *
1003) Force.Ratio>=0.4477727 38 12 Defender Wins (0.3157895 0.6842105) *
251) Tank.Ratio>=0.8546171 9 1 Defender Wins (0.1111111 0.8888889) *
63) Initial.Force.Ratio>=1.268086 13 0 Defender Wins (0.0000000 1.0000000)
*

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B.

1. **MODEL 2.1**
    1) root 27 9 Attacker Wins (0.66666667 0.33333333)
    2) LEADA>=0.5 15 1 Attacker Wins (0.93333333 0.06666667) *
    3) LEADA< 0.5 12 4 Defender Wins (0.33333333 0.66666667) *

2. **MODEL 2.2**

    1) root 40 15 Attacker Wins (0.6250000 0.3750000)
    2) LEADA>=-0.5 32 7 Attacker Wins (0.7812500 0.2187500)
    3) Arty.Ratio>=1.015957 12 0 Attacker Wins (1.0000000 0.0000000) *
    4) Arty.Ratio< 1.015957 20 7 Attacker Wins (0.6500000 0.3500000)
    10) LEADA>=0.5 11 2 Attacker Wins (0.8181818 0.1818182) *
    11) LEADA< 0.5 9 4 Defender Wins (0.4444444 0.5555556) *
    3) LEADA< -0.5 8 0 Defender Wins (0.0000000 1.0000000) *
3. **Model 2.3**
    1) root 76 35 Defender Wins (0.46052632 0.53947368)
    2) LEADA>=0.5 26 1 Attacker Wins (0.96153846 0.03846154) *
    3) LEADA< 0.5 50 10 Defender Wins (0.20000000 0.80000000)
    6) Cav.Ratio< 0.6976321 9 3 Attacker Wins (0.66666667 0.33333333) *
    7) Cav.Ratio>=0.6976321 41 4 Defender Wins (0.09756098 0.90243902) *

4. **Model 2.4**
    1) root 95 43 Attacker Wins (0.5473684 0.4526316)
    2) INITA>=0.5 50 10 Attacker Wins (0.8000000 0.2000000) *
    3) INITA< 0.5 45 12 Defender Wins (0.2666667 0.7333333)
    6) Force.Ratio>=3.363226 8 3 Attacker Wins (0.6250000 0.3750000) *
    7) Force.Ratio< 3.363226 37 7 Defender Wins (0.1891892 0.8108108) *
5. **Model 2.5**
    1) root 133 51 Attacker Wins (0.6165414 0.3834586)
    2) INITA>=0.5 90 22 Attacker Wins (0.7555556 0.2444444)
    3) AEROA>=-0.5 72 11 Attacker Wins (0.8472222 0.1527778) *
    4) AEROA< -0.5 18 7 Defender Wins (0.3888889 0.6111111) *
    3) INITA< 0.5 43 14 Defender Wins (0.3255814 0.6744186)
    6) Arty.Ratio>=2.221951 19 9 Attacker Wins (0.5263158 0.4736842) *
    7) Arty.Ratio< 2.221951 24 4 Defender Wins (0.1666667 0.8333333) *
6. **Model 2.6**
    1) root 51 18 Attacker Wins (0.6470588 0.3529412)
    2) AEROA>=0.5 37 6 Attacker Wins (0.8378378 0.1621622) *
    3) AEROA< 0.5 14 2 Defender Wins (0.1428571 0.8571429) *

7. Model 2.7

1) root 618 217 Attacker Wins (0.64886731 0.35113269)
2) INITA>=0.5 393 82 Attacker Wins (0.79134860 0.20865140)
3) LEADA>=-0.5 362 56 Attacker Wins (0.84530387 0.15469613)
8) LEADA>=0.5 143 4 Attacker Wins (0.97202797 0.02797203) *
9) LEADA< 0.5 219 52 Attacker Wins (0.76255708 0.23744292)
18) Force.Ratio>=1.08765 182 27 Attacker Wins (0.85164835 0.14835165) *
19) Force.Ratio< 1.08765 37 12 Defender Wins (0.32432432 0.67567568) *
5) LEADA< -0.5 31 5 Defender Wins (0.16129032 0.83870968) *
3) INITA< 0.5 225 90 Defender Wins (0.40000000 0.60000000)
6) Close.Air.Support.Ratio>=0.9775281 70 23 Attacker Wins (0.67142857 0.32857143)
12) MOMNTA< -5 25 3 Attacker Wins (0.88000000 0.12000000) *
13) MOMNTA>=-5 45 20 Attacker Wins (0.55555556 0.44444444)
26) Force.Ratio>=3.894631 14 1 Attacker Wins (0.92857143 0.07142857) *
27) Force.Ratio< 3.894631 31 12 Defender Wins (0.38709677 0.61290323) *
7) Close.Air.Support.Ratio< 0.9775281 155 43 Defender Wins (0.27741935 0.72258065)
14) LEADA>=0.5 18 5 Attacker Wins (0.72222222 0.27777778) *
15) LEADA< 0.5 137 30 Defender Wins (0.21897810 0.78102190) *

APPENDIX C.

1. MODEL 3.1

1) root 27 9 Attacker Wins (0.66666667 0.33333333)
2) LEADA>=0.5 15 1 Attacker Wins (0.93333333 0.06666667) *
3) LEADA< 0.5 12 4 Defender Wins (0.33333333 0.66666667) *

2. MODEL 3.2

1) root 45 17 Attacker Wins (0.6222222 0.3777778)
2) LEADA>=-0.5 36 8 Attacker Wins (0.7777778 0.2222222)
3) TERRA1=RM0,RW0 30 4 Attacker Wins (0.8666667 0.1333333) *
4) TERRA1=00U,FM0,RB0 6 2 Defender Wins (0.3333333 0.6666667) *
5) LEADA< -0.5 9 0 Defender Wins (0.0000000 1.0000000) *

3. MODEL 3.3

1) root 76 35 Defender Wins (0.46052632 0.53947368)
2) LEADA>=0.5 26 1 Attacker Wins (0.96153846 0.03846154) *
3) LEADA< 0.5 50 10 Defender Wins (0.20000000 0.80000000)
4) Cav.Ratio< 0.6976321 9 3 Attacker Wins (0.66666667 0.33333333) *
5) Cav.Ratio>=0.6976321 41 4 Defender Wins (0.09756098 0.90243902) *

4. Model 3.4

1) root 95 43 Attacker Wins (0.54736842 0.45263158)
2) INITA>=0.5 50 10 Attacker Wins (0.80000000 0.20000000) *
3) INITA< 0.5 45 12 Defender Wins (0.26666667 0.73333333)
4) WX1=DOT$T,DSTST,DSTWT,WLCFT 21 10 Attacker Wins (0.52380952 0.47619048)
12) Force.Ratio>=1.657281 8 2 Attacker Wins (0.75000000 0.25000000) *
13) Force.Ratio< 1.657281 13 5 Defender Wins (0.38461538 0.61538462) *
5) WX1=DOTST,DOTWT,DSCFT,DSHST,DST$T,DSTFT,WHC$T,WHT$T,WHTST,WLT$T,WLTFT,WOTST 24 1 Defender Wins (0.04166667 0.95833333) *

5. Model 3.5

1) root 133 51 Attacker Wins (0.6165414 0.3834586)
2) WX1=DOCFT,DOTFT,DOTST,DOTWT,DSCWD,DSCWT,DST$T,DSTFT,DSTST,WHCFT,WHCWD,WHCWT,WHHWE,WHTFT,WHTWT,WLCFT,WLT$T,WLTST 96 24 Attacker Wins (0.7500000 0.2500000)
3) AEROA>=-0.5 81 14 Attacker Wins (0.8271605 0.1728395) *
4) AEROA< -0.5 15 5 Defender Wins (0.3333333 0.6666667) *

3)WX1=D0CWD,D0T$T,DSCFT,DSHST,DSTWT,W0CWT,WLC$T,WLC WT,WLTF T,WLTWT 37 10 Defender Wins (0.2702703 0.7297297)

6) Initial.Force.Ratio>=1.453048 24 10 Defender Wins (0.4166667 0.5833333)

12) POST1=HD,PD 10 3 Attacker Wins (0.7000000 0.3000000) *

13) POST1=DL,FD 14 3 Defender Wins (0.2142857 0.7857143) *

7) Initial.Force.Ratio< 1.453048 13 0 Defender Wins (0.0000000 1.0000000) *

6. <u>Model 3.6</u>

1) root 51 18 Attacker Wins (0.64705882 0.35294118)
2) WX1=D0HFD,D0HSD,D0T$E,D0T$T,D0TWT,DSHST,WLT$T,WLTST 35 3 Attacker Wins (0.91428571 0.08571429)
3) Initial.Force.Ratio>=-1750 32 1 Attacker Wins (0.96875000 0.03125000) *
4) Initial.Force.Ratio< -1750 3 1 Defender Wins (0.33333333 0.66666667)
5) Close.Air.Support.Ratio>=26 1 0 Attacker Wins (1.00000000 0.00000000) *
6) Close.Air.Support.Ratio< 26 2 0 Defender Wins (0.00000000 1.00000000) *

3) WX1=DSHFT,DSTFT,WHTST 16 1 Defender Wins (0.06250000 0.93750000)

6) Force.Ratio>=6.243036 1 0 Attacker Wins (1.00000000 0.00000000) *

7) Force.Ratio< 6.243036 15 0 Defender Wins (0.00000000 1.00000000) *

7. <u>Model 3.7</u>

1) root 618 217 Attacker Wins (0.64886731 0.35113269)
2) INITA>=0.5 393 82 Attacker Wins (0.79134860 0.20865140)
3) LEADA>=-0.5 362 56 Attacker Wins (0.84530387 0.15469613)

8) LEADA>=0.5 143 4 Attacker Wins (0.97202797 0.02797203) *

9) LEADA< 0.5 219 52 Attacker Wins (0.76255708 0.23744292)

18) Force.Ratio>=1.08765 182 27 Attacker Wins (0.85164835 0.14835165) *

19) Force.Ratio< 1.08765 37 12 Defender Wins (0.32432432 0.67567568)

38) WX1=D0CWT,DSHST,DSTFT,WHCWT,WLCFT 10 3 Attacker Wins (0.70000000 0.30000000) *

39) WX1=DOTST,DSCWT,DST$T,DSTST,WLCWT,WLTST 27 5 Defender Wins (0.18518519 0.81481481) *

5) LEADA< -0.5 31 5 Defender Wins (0.16129032 0.83870968) *

3) INITA< 0.5 225 90 Defender Wins (0.40000000 0.60000000)

6) WX1=D0HFD,D0HSD,D0T$D,D0T$E,D0T$T,D0TFT,D0TWD,D0TWT,DOCFT,DOT$T,WHCWT,WHTFT,WLCFT,WLTST 35 3 Attacker Wins (0.91428571 0.08571429) *

7) WX1=D0CWD,DOTST,DOTWT,DSCFT,DSH$T,DSHFT,DSHST,DST$T,DSTFT,DSTST,DSTWT,W0CWT,WHC$T,WHCFT,WHT$T,WHTSE,WHTS

T,WHTWT,WLCT,WLTT,WLTFT,WLTWT,WOTST 190 58 Defender Wins (0.30526316 0.69473684)
14) Force.Ratio>=4.074323 23 7 Attacker Wins (0.69565217 0.30434783)
28) WX1=DSHFT,DST$T,DSTFT,DSTST,DSTWT,WLTFT 13 0 Attacker Wins (1.00000000 0.00000000) *
29) WX1=DSH$T,DSHST,W0CWT,WHT$T,WHTSE 10 3 Defender Wins (0.30000000 0.70000000) *
15) Force.Ratio< 4.074323 167 42 Defender Wins (0.25149701 0.74850299)
30) LEADA>=0.5 19 6 Attacker Wins (0.68421053 0.31578947) *
31) LEADA< 0.5 148 29 Defender Wins (0.19594595 0.80405405) *

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Arquilla John and David Ronfeldt. *Networks and Netwars: The Future of Terror, Crime, and Militancy.* Santa Monica, CA: Rand Corporation, 2001.

Aguilar, GC. D. Napoleon's Maxims of War with notes by General Burnod. *Pennsylvania,* Pa: David McKay, 1902.

BatchGeo, L. L. C. "BatchGeo database." http://www.batchgeo.com,

Bracken, Jerome. "Lanchester Models of the Ardennes Campaign." *Naval Research Logistic* 42, no. 4 (1995): 559–577.

Breiman Leo, Jerome Friedman, Charles J. Stone, and Richard A. Olshen. *Classification and Regression Trees.* Boca Raton, FL: CRC Press, 1984.

Chandler David P. *The Land and People of Cambodia.* New York, NY: HarperCollins Publishers, 1991.

Clemens, SC. "The Application of Lanchester Models to the Battle of Kursk." *Unpublished Manuscript, Yale University* 5, (1997).

Coban Muzaffer. "Predicting Battle Outcomes with Classification Trees." Master's thesis, Naval Postgraduate School, 2001.

Collins, Joseph J. "The Soviet Invasion of Afghanistan: A Study in the use of Force in Soviet" *Foreign Policy,* 1986.

Coolahan, E. James. "A Vision for Modeling and Simulation at APL." *Johns Hopkins APL Technical Digest* 26, no. 4 (2005).

Crawley, Michael J. *The R Book.* Hoboken, NJ: John Wiley & Sons, 2012.

Croicu Mihai Cătălin, Stina Högbladh, Therése Pettersson, and Lotta Themnér. "UCDP External Support Project Primary Warring Party Dataset Codebook." (2011): 1–2011.

Department of Defense. *Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A), Instructions Number 5000.61.* Washington, D.C.:2009.

Department of Defense. *Irregular Warfare Joint Operating Concept (JOC).* Washington, D.C.: 2007.

Department of the Army. *Army Special Operations Forces Unconventional Warfare* (FM 3–05.130). Washington, D.C.: Army Publishing Directorate, 2008.

Department of the Army. *Special Forces Operations* (FM 31 -20). Washington, D.C.: Army Publishing Directorate, 2001.

Deutsch, Morton, Peter T. Coleman, and Eric C. Marcus. *The Handbook of Conflict Resolution: Theory and Practice*. Hoboken, NJ: John Wiley & Sons, 2011.

DeVotta, Neil. "The Liberation Tigers of Tamil Eelam and The Lost Quest for Separatism in Sri Lanka." Asian Survey (2009): 1021–1051.

Dupuy, Trevor Nevitt. *Numbers, predictions, and war: Using History to Evaluate Combat Factors and Predict the Outcome of Battles*. Virginia: NOVA Publications 1985.

Earle, Edward Mead, Gordon Alexander Craig, and Felix Gilbert. *Makers of Modern Strategy: Military thought from Machiavelli to Hitler*. Princeton, NJ: University Press, 1943.

Eck, Kristine and Lisa Hultman. "One-Sided Violence Against Civilians in War Insights from New Fatality Data." *Journal of Peace Research* 44, no. 2 (2007): 233–246.

Faraway, Julian J. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Boca Raton, FL: CRC press, 2005.

Fricker, Ronald D. "Attrition Models of the Ardennes Campaign." *Naval Research Logistics* 45, no. 1 (1998): 1–22.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. "Additive Logistic Regression: A Statistical View of Boosting." *The Annals of Statistics* 28, no. 2 (2000): 337–407.

Fuller, John Frederick Charles. *The Conduct of War, 1789–1961: A Study of the Impact of the French, Industrial, and Russian Revolutions on War and its Conduct*. Boston, MA: Da Capo Press, 1992.

Gat, Azar. *War in Human Civilization*. New York, NY: Oxford University Press, 2006.

Gunn, Geoffrey C. *Political struggles in Laos, 1930–1954: Vietnamese Communist Power and the Lao Struggle for National Independence*. Bangkok, Thailand: Editions Duang Kamol, 1988.

Halperin, Morton H. "Guaranteeing Democracy." *Foreign Policy* (1993): 105–122.

Hammes, Thomas X. "The Evolution of War: The Fourth Generation." *Marine Corps Gazette* 78, no. 9 (1994): 35–44.

Hammes Thomas X. "War Evolves into the Fourth Generation." *Contemporary Security Policy* 26, no. 2 (2005): 189–221.

Hansen Matthew, R. Dubayah, and R. DeFries. “Classification Trees: An Alternative to Traditional Land Cover Classifiers.” *International Journal of Remote Sensing* 17, no. 5 (1996): 1075–1081.

Harbom Lotta. “UCDP Dyadic Dataset Codebook.” *Uppsala Conflict Data* (2009).

Harbom Lotta, Stina Högbladh, and Peter Wallensteen. “Armed Conflict and Peace Agreements.” *Journal of Peace Research* 43, no. 5 (2006): 617–631.

Hartley Dean S. and Robert L. Helmbold, “Validating Lanchester’s Square Law and Other Attrition Models.” *Naval Research Logistics (NRL)* 42, no. 4 (1995): 609–633.

Helmbold, Robert L. *Rates of Advance in Historical Land Combat Operations*. (CAA-RP-90-1). Bethesda, MD: U.S. Army Concepts Analysis Agency, 1990.

Helmbold, Robert L. and Aqeel A. Khan, *Combat History Analysis Study Effort (CHASE)*(CAA-TP-86-2). Bethesda, MD: U.S. Army Concepts Analysis Agency, 1986.

Howard, Michael. “The Causes of Wars.” *The Wilson Quarterly* (1984): 90–103.

Howard, Michael.”War and the Nation-State.” *Daedalus* 108, no.4 (1979): 101–110.

JMP, A. and Marcel Proust. “Specialized Models.” Cary, NC: SAS (2013).

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. “The Worldwide Governance Indicators: Methodology and Analytical Issues.” *Hague Journal on the Rule of Law* 3, no. 02 (2011): 220–246.

Kreutz, Joakim. “How and When Armed Conflicts End: Introducing the UCDP Conflict Termination Dataset.” *Journal of Peace Research* 47, no. 2 (2010): 243–250.

Larson, Eric V., Derek Eaton, Brian Nichiporuk, and Thomas S. Szayna. *Assessing Irregular Warfare: A Framework for Intelligence Analysis*. Santa Monica, CA: Rand Corporation, 2009.

Lind, William S. “Understanding Fourth Generation Warfare.” *Military Review* no. 5 (2004): 12–16.

Lucas, Thomas W. and John A. Dinges. “The Effect of Battle Circumstances on Fitting Lanchester Equations to the Battle of Kursk.” *Military Operations Research* 9, no. 2 (2004): 17–30.

Lucas Thomas W and Turker Turkes. “Fitting Lanchester Equations to the Battles of Kursk and Ardennes.” *Naval Research Logistics (NRL)* 51, no. 1 (2004): 95–116.

Macal M. Charles and Michael J. North. “Tutorial on Agent-based Modelling and Simulation.” *Journal of Simulation* 4, no. 3 (2010): 151–162.

Martel, C. William. *Victory in War: Foundations of Modern Strategy*. Vol. Revised and Expanded Edition. New York, NY: Cambridge University Press, 2011.

McQuine Robert. *Historical Characteristics of Combat for Wargames*. Bethesda, MD: U.S. Army Concepts Analysis Agency, 1988.

Melander Erik, Frida Möller, and Magnus Öberg. “Managing Intrastate Low-Intensity Armed Conflict 1993–2004: A New Dataset.” *International Interactions* 35, no. 1 (2009): 58–85.

Melander Erik and Nina von Uexkull. “Sustained Third Party Engagement and Conflict Termination: An Introduction of the UCDP Managing Intrastate Conflict (MIC) Dataset.”2011.

Orend Brian. *The Morality of War*. Calgary, Canada: Broadview Press, 2013.

Paret Peter. “Clausewitz in His Time: Essays in the Cultural and Intellectual History of Thinking about War.” New York, NY: Berghahn Books, 2014.

Kretchik, Walter Edward. *Invasion, Intervention,” Intervasion”: A Concise History of the U.S. Army in Operation Uphold Democracy*. Darby, PA: DIANE Publishing, 1998.

Perry Nigel. *Applications of Historical Analyses in Combat Modelling*. Joint Operations Division, Defense Science and Technology Organization, Canberra, Australia: 2011.

Policy, Foreign. “Fund for Peace (2006).” *Failed States Index*.

“Port-Au-Prince, Haiti.” World Weather and Climate Information. http://www.weather-and-climate.com/average-monthly-Rainfall-Temperature-Sunshine,Port-Au-Prince, Haiti.

Pro, JMP. “v10.” *Cary, NC: SAS* (2012).

Raleigh Clionadh and Caitriona Dowd. “Armed Conflict and Event Data Project (ACLED) Codebook.” (2015).

Raleigh Clionadh, Andrew Linke, Havard Hegre and Joakim Karlsen Karlsen. “Armed Conflict Location and Event Data.” Journal of Peace Research 47(5) 1–102015.

Raleigh Clionadh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. “Introducing ACLED: An Armed Conflict Location and Event Dataset Special Data Feature.” *Journal of Peace Research* 47, no. 5 (2010): 651–660.

Sarkees Meredith Reid. “The COW Typology of War: Defining and Categorizing Wars (Version 4 of the Data).” *Note with Version 4 of the Correlates of War Data* (2010).

Sarkees Meredith Reid. “Non-State Wars (Version 4.0).” *Definitions and Variables. Correlates of War Project* (2010).

Singer, Joel David and Melvin Small. *The Wages of War, 1816–1965: A Statistical Handbook* . Hoboken, NJ: John Wiley & Sons, 1972.

Strong, Carol and Gillum, Joshua. “Unraveling the Gordian Knot of Strategic U.S. Military Engagement: Asymmetrical Relationships, Unconventional Means and International Conflict.” Paper prepared for Cultural Diplomacy and International Relations: New Actors, New Initiatives and New Targets Institute for Cultural Diplomacy, Annual Conference, Berlin, Germany 15–18 November, 2011.

Sundberg, Ralph. “Revisiting One-Sided Violence–a Global and Regional Analysis.” *States in Armed Conflict, Ed.Lotta Harbom and Ralph Sundberg.Uppsala: Universitetstryckeriet* (2009).

Sundberg, Ralph, Kristine Eck, and Joakim Kreutz. “Introducing the UCDP Non-State Conflict Dataset.” *Journal of Peace Research* 49, no. 2 (2012): 351–362.

Team, R. Core. *R Language Definition* (2000).

Themnér, Lotta and Peter Wallensteen. “Armed Conflicts, 1946–2011.” *Journal of Peace Research* 49, no. 4 (2012): 565–575.

Trevor N. Dupuy. The Dupuy Institute2015. http://www.dupuyinstitute.org/tndupuy.htm.

Tolk, Andreas. *Engineering Principles of Combat Modeling and Distributed*. Hoboken, NJ: John Wiley & Sons, 2012.

Un, Kheang. *Democratization without Consolidation: The Case of Cambodia, 1993--2004*. DeKalb, IL: Northern Illinois University, 2004.

Uppsala Conflict Data Program, Uppsala University. “UCDP Database Categorical Variables 1989–2008.” (1989-2008).

Vego, Milan. “German War Gaming.” *Naval War College Review* 65, no. 4 (2012).

Wonnacott, Thomas H., and Ronald J. Wonnacott. *Introductory Statistics*. Vol. 19690. New York, NY: John Wiley& Sons, 1972.

World Bank Group, ed. *World Development Indicators 2012*. Washington, D.C.: World Bank Publications, 2012.

Yigit Faruk. "*Finding the Important Factors in Battle Outcomes: A Statistical Exploration of Data from Major Battles.*" Master's thesis, Naval Postgraduate School, 2000.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California